

Simulation of Consciousness, and pragmatic accounting for language

Dr. Stephen Rainey¹

Abstract

The possibilities for the simulation of consciousness are convincingly critiqued by Chinese Room type arguments. This paper discusses what meaning the simulation of consciousness could have anyway, in terms of objectivity and subjectivity. In terms of an attempt to create *in silico* (or otherwise) a simulation of consciousness this means seeing how ascriptions of such terms ought to be made. Drawing upon Sellars, Kant, and Merleau-Ponty, it is argued that whether or not something *really* manifests consciousness is not the only question worth pursuing (firstly needing clarification). Further, it suggests that it is plausible that pragmatic accounts of meaningful language use could pose novel questions for Chinese Room type arguments in moving from syntax-plus-semantic accounts of language to an account based in the deployment of abilities.

Introduction

Chinese Room type arguments can be framed as successfully problematizing claims to subjectivity for the system by denying syntax is sufficient for semantics [1], [2] – there is no meaning to the system’s activity so no relevant subjectivity in it. However, the question remains as to the objectivity of subjectivity [3]. Where meaning and subjectivity are linked it is worth asking how, and to what extent, what kinds of things can contribute to that meaning – including what objects.

The brain is a very interesting object not least owing to the apparent role it plays in the manifestation of consciousness. Somehow, the brain seems to be involved in whatever it takes to think, perceive, feel emotions, and generally to be a locus for uniquely conscious perspectives. Some may take it that the brain itself is a structure that is intrinsically conscious [4], [5]. Others may reject that, suggesting the brain has a role in a more distributed account of consciousness, maybe one crossing the brain/body/world boundary or even one that connects to a non-extended world of thought [6], [7].

The Human Brain Project (HBP) is a European research initiative that aims to integrate neuroscientific research in order to “...advance neuroscience and medicine and to create brain-inspired information technology.” [8, p. 574] This motivates an analysis of cognition and computation, and the technical possibilities of their interplay.

Understanding the brain

Where ‘understanding the human brain’ is an aim, the idea of looking at consciousness most likely arises. This is because the

brain appears to be related tightly to conscious activity and processes. This is salient as the majority of objects routinely encountered in the universe do not appear to relate to consciousness in the way the brain does. It is very widely proposed that the brain is at least somehow, if not exhaustively, responsible for conscious phenomena. This presents a subject/object dichotomy that can be explored with reference to Wilfrid Sellars.

The Manifest and the Scientific Image

There are at least two ways to account for how human beings can conceive of and experience the world around them. One of Wilfrid Sellars’ main preoccupations concerns this apparently dual nature of human experience and knowledge. This dual nature is characterised by the ‘manifest image’ and the ‘scientific image’. The former is a set of experiences and knowledge concerning how it is to be a person in a world of experience – lessons from living. The latter is a set of insights into the workings of the world from an experimental point of view – lessons from physics, chemistry and so on. Both images appear to be well founded, and yet they are in tension with one another in important ways [9].

The manifest image is seemingly immediate, objective, complete and real. However, being knowledgeable about scientific theories and discoveries, it is also apparent that the object before us is part of a system of physical entities with properties and interactions themselves not apparent. The scientific image lets us know that the experience we have can be accounted for differently, and not incorrectly, especially by reducing it to various sub-properties. Colour is the reflection and absorption of light by surfaces, for example. These reductive, sub-properties aren’t experienced as such, yet they stand in a system of explanations for the experience, as if explaining the obvious in terms of the inscrutable [10].

Two of the tensions between the images can be labelled in general as tensions of authenticity, and tensions of instrumentality. Authenticity of experience is diminished in the scientific image from the perspective of the manifest. Instrumentality is diminished in the manifest image from the perspective of the scientific. So a reductive account of the colour pink, in terms of wavelengths absorbed and reflected by plane surfaces, misses what makes the manifest experience of pink an experience of pink – it seems inauthentic. An authentic account of the manifest experience of pink gives nothing to the scientific image to work with – ‘it is pink, what now?’ could be the question here. This sketch of Sellars’ perspective here perhaps suggests that there is considerable difficulty to be found in resolving the manifest and scientific images. Effectively, prior to much reflection and theorising, the world appears stereoscopically – there are subjects and there are objects. A synoptic vision, if possible, is not apparent in experience.

Simulation and Sellars

Is it the manifest image of the other that ought to be re-created in technical projects, or the scientific? This question seems pressing where, for instance, artificial intelligence is sought. Do such endeavours aim to re-create the manifest image of being a mind in the world, or to re-create sets of processes and properties

¹ De Montfort University, Leicester, Centre for Computing and Social Responsibility

constituting a scientific image of a mind (whatever this might be)?

To create the manifest mind we would need to create something that can be taken for a mind. To create the scientific mind we would need to create something that reproduced processes and properties that enabled a mind to develop. In either case, what a mind *is* must be known. This is no simple question.

On an internalist account, consciousness is produced somehow by the intrinsic structures and properties of the brain itself – its structures, the material from which it is hewn etc. If thought of in an externalist way, consciousness arises not exclusively from the brain, but by worldly relations in which the brain, its structures and properties find themselves. In either case, the brain is important in an account of consciousness. Understanding the brain, therefore, should have implications for understanding consciousness – perhaps minimally by weighing in on the internalist / externalist debate. This might also be stated the other way around: perhaps without an account of consciousness, the brain can't properly be understood in the first place. The role for prior theory is illustrated next.

From Crick and Koch [4] it is hypothesised that a key element in understanding consciousness in human beings at least comes from the synchronous oscillations of the brain during visual stimulation. These occur in the 40-70Hz range during visual experience and are held to be explanatory of the binding of visual stimulations into a visual experience. In looking at synchrony as key for consciousness, Thompson and Varela [6, p. 418] argue for the essential importance of environment in conjunction with the body, and its states in a 'seamless' way. Clark [11] meanwhile, concludes that only an internalist account of consciousness could explain the speed at which the brain activity occurs. The seamlessness of any relation still could not provide the necessary proximities for the time-scales observable. Which kind of account, internalist or externalist, should we choose to account for conscious experience here? Given these interpretations of the data, it seems an open question. In the first instance, the complex dynamism of an organism's embeddedness in an environment prompts the conclusion that conscious experience must be thought of as extending beyond the confines of the head. In the second, though, the detail of the conscious experience are held to prompt the conclusion that it can only be explained as an internalist phenomenon.

A synoptic vision

Between Sellars, and the equivocal nature of evidence depending on prior theory, what can be said of the objective reality of subjectivity? This is a key question especially where technology would aim to re-create something like a human other, a locus of subjectivity. As [12] suggest, using non-humans in order to gain insight into humans is not only possible, but can be acutely useful in particular cases, e.g. "by implementing neurocognitive mechanisms in androids and then evaluating their interactions with people, androids have the potential to help researchers bridge the gap between cognitive neuroscience and the behavioral sciences, leading to a new way of understanding individuals." [12, p. 321]

Here, perhaps surprisingly, Merleau-Ponty has an interesting contribution. If we ask ourselves what makes a self a self at all,

"I say that it is another, a second self, and this I know in the first place because this living body has the same structure as mine. I experience my own body as the power of adopting certain forms of behaviour and a certain world, and I am given to myself merely as a certain hold upon the world; now, it is precisely my body which perceives the body of another, and discovers in that other body a miraculous prolongation of my own intentions, a familiar way of dealing with the world. Henceforth, as the parts of my body together comprise a system, so my body and the other's are one whole, two sides of one and the same phenomenon, and the anonymous existence of which my body is the ever-renewed trace henceforth inhabits both bodies simultaneously." [3, p. 412]

In this can be read a summary of the reality of objectivity from a Kantian point of view, along with the necessity of a unity of apperception – this description seeks to elucidate the dual nature of purportedly individual experience. The interesting contribution here is that subjectivity, or intentionality, is transformed from what might be assumed to have an essential nature. Merleau-Ponty instead presents it as anonymous and shared. This anonymous shared subjectivity revealed in the objective apparent 'prolongation of one's own intentions'.

The reality of subjectivity is revealed in the otherness of the objective agent with whom I am presented. The 'hold' the subject has on the world is recognised as the apparent exercise of agency in objectivity to the extent that this can be seen emanating from the other. The objectivity of a (minimally) binary 'us' revealed in the moment of recognition in action convinces oneself of oneself. To the extent that it succeeds in describing the objectivity of subjectivity, it re-draws the problematic from Sellars but with a Kantian optimism – the idea that one can only act at all in the unprovable faith that one's action can affect a transcendent reality – positing a synoptic vision of scientific and manifest image. Experience in the immediate sense is complemented at the same time in a question about the other and their position as significantly like, yet other than, oneself.

This can be taken as a way of grounding an appreciation of cognition, of consciousness, of intelligence, in the simulation of behaviour. It requires an ethical position that is neutral between subjects and objects in terms of potential ascription of worth. It requires that the world of experience be seen as a space of reasons, in which existential as well as descriptive status are matters of discourse not essence.

Pragmatic accounting for language

Robert Brandom accounts for human linguistic ability in terms of human beings as 'concept-mongers'. This means that human beings have an aptitude for operating in a world of reasons, which manifests in practical abilities around inference. Implicit in much of human behaviour are such practical inferences, and these form a basis for concluding that human behaviours, such as communicative use of language, are meaningful.

Where an account of consciousness is pursued along the lines drawn above – in terms of discourse, evaluative, and on matters existential and description – might novel questions arise for Chinese Room type arguments? These arguments succeed in showing the inability of a syntax engine to manifest semantic content, hence the impossibility of machine consciousness. But

pragmatism takes a world relation from action to meaning – not the manifestation of meaning from a formal basis.

Where a ‘hold’ upon us can manifest through the recognition of anonymous second self-hood where we can apprehend action, might this not extend to matters of understanding? Communicative language is after all *action*.

Especially where an android type scenario is envisioned, it is plausible that *prima facie* speech actions could manifest such that they at least appear to be meaningful. It could thus be argued that, from a pragmatic account of language, meaningful abilities could be simulated so as to manifest apparent conscious acts. Where consciousness is accounted for as anonymous and shared, appearances may be all that is required, as such appearances are sufficient to prompt, or to sustain, a discourse on the very idea of consciousness. Where consciousness isn’t possessed by the conscious agent, but is partaken in as a kind of questioning, this is an issue worthy of further exploration.

Grasp of a concept is an aptitude to deploy inferential abilities, including giving and asking for reasons. [13, p. 89] This is because any number of justifications for one utterance or another is possible: the inclusion of particular inferential transitions grounding, or accounting for, some utterance is contingent upon the particular meaning being deployed and is in principle not knowable in a complete sense. Giving offence, for instance, through deploying a term like ‘Irish’ to mean perplexingly obtuse, or daft [14, p. 327], is not contained within the concept ‘Irish’ but it can be so deployed where the will exists to shore it up with particular explanations [15].

More resources for evaluation are required, this being the case, when it comes to assessing the appearance or not of meaningful conceptual deployment. Validity of conceptual inference depends on more than propositional calculus. It requires understanding of context, which in turn includes a practical ability to assess practical abilities.

Non verbal ‘justification’

‘Knowing how to go on’ is one practical way in which one could be said to be justifying one’s action in some context. Raising my hand at an oncoming bus lets the driver know I want to board, and demonstrates I know how to catch a bus. Improvising a piece of music demonstrates my knowledge of key, rhythm, tone, and makes a tune. It might be that the majority of ‘know how’ put to work in a day is this kind of tacit, non-verbally explained sort [16].

Example: Karate Roboy

For context, we are to imagine a human who is taught karate, but quite obliquely.[17] A master of the art takes on the student only to confound them by making them carry out chores such as painting the fence, waxing the car, and sweeping the floor. The master is fastidious about how the movements must be carried out, to the near-exasperation of the student. Upon almost having enough, the student finally protests, only for the master to reveal that these movements constituting the chores were exactly those required to carry out essential karate moves. This is demonstrated to the student by demanding the chore movements, which are seen to be sufficient for deflecting punches, kicks etc. Without really knowing it, by learning how to do chores, the student has learned how to do karate. From the context of chores,

with only a change of context into that of being attacked, the student’s movements are no longer ‘painting the fence’, but ‘blocking a strike’, and so on. The student knows karate, having learned chores.

Now we transmute the scenario into a robot of a particular kind. A non-von Neumann architecture is present in an embodied robot; the control centres of the machine are neuromorphic (with neuron firing, synapse-like modifications etc.).[18] The embodiment is in a “Roboy”, a specific kind of robotic platform predicated on smooth actuation, based on a bio-mimetic approach.[19]

Roboy learns chore-movements by taking in the visual stimuli, of chores being carried out. These are not ‘seen as’ chores, simply as dynamic orientations in space. Roboy can mimic the movements well. It just so happens that these movements – painting a fence, waxing a car, sweeping the floor – are identical with those of some key karate moves – deflections, strikes etc. The chore movements are all Roboy has ‘learned’. Suddenly, Roboy finds itself in a karate scenario – nefarious bots from a rival lab appear and are programmed to strike. If Roboy does what it can, deploys its abilities in the new context, is it now doing karate or merely doing chore-like movements out of context? How would we assess this question?

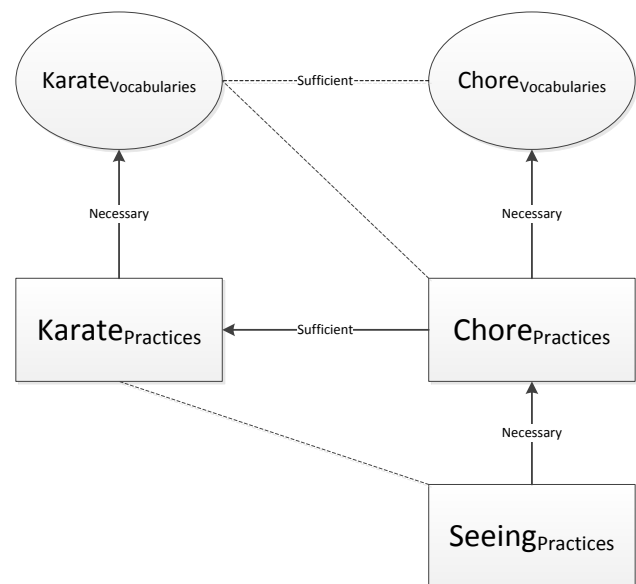


Figure 1 Practices and vocabularies

In the above figure, we see a depiction of how the practices of carrying out chores and necessary for the ability to discuss how chores ought to be done – Chore_{vocabularies}. We see that it transpires Karate_{practices} are grounded sufficiently in the same movements. The karate practices are necessary for the vocabularies of karate, which are in turn sufficient for those of chore vocabularies. The relation between chore practices and karate vocabularies is not clear. Neither is it clear, and this is the area for discussion, how Roboy’s watching of Chore_{practices} links with karate practices whereby the nature of the thought experiment Roboy does no ‘seeing as’.

MIT researchers have developed a neural net that can deploy a justification for output [20]. Does Roboy's processing constitute such a justification? The neuromorphic processing of movement in space produces a readout that could be taken as a justification of why this movement now, given that movement then. What is the difference between this kind of sufficiency and a meaningful deployment of conceptual ability?

More besides a descriptive account of an action is required, it seems, and that relies on the manifest image account of things. This is presumably missing both for the MIT moral machine, and for Roboy's karate: there is no account of the propriety of this or that action, there is just an account of this or that action. Nevertheless, there appears to be an interesting case for pragmatic accounts of language in dealing with simulation of behaviour and genuine action.

Conclusion

Simulation of behaviour might be capable of prompting in the actually conscious a realisation of the objectivity of subjectivity. It might not be subjectivity, however. The content of pragmatic accounting for language suggests 'knowing how to proceed' could be a marker of conceptual grasp, and so perhaps of meaningfulness. Combined with the apparent objective manifestation of anonymous subjectivity it might be seductive to conclude meaningful subjectivity can be created and evidence in action. But without the proper relations among actions – including a grasp of the propriety among actions – this is only a superficial suggestion.

Roboy learning karate via copying chores won't demonstrate a counter-example to Chinese Room objections to artificial mental states, but it might be a warning that kung-fu zombie robots are a near-future possibility.

Acknowledgement

This research has received funding from the European Union Horizon 2020 framework programme and grant agreement number 720270 (HBP SGA1). The views represented are those of the authors, not the Human Brain Project as a whole.

References

- [1] J. R. Searle, 'Minds, Brains and Programs', *Behav. Brain Sci.*, vol. 3, no. 3, pp. 417–57, 1980.
- [2] S. J. Nasuto, J. M. Bishop, E. B. Roesch, and M. C. Spencer, 'Zombie Mouse in a Chinese Room', *Philos. Technol.*, vol. 28, no. 2, pp. 209–223, Jun. 2015.
- [3] M. Merleau-Ponty, *Phenomenology of Perception*. Routledge, 2002.
- [4] F. Crick and C. Koch, 'Towards a neurobiological theory of consciousness', in *Seminars in the Neurosciences*, 1990, vol. 2, pp. 263–275.
- [5] J. Prinz, *The Conscious Brain*. OUP USA, 2012.
- [6] E. Thompson and F. J. Varela, 'Radical embodiment: neural dynamics and consciousness', *Trends Cogn. Sci.*, vol. 5, no. 10, pp. 418–425, 2001.
- [7] S. Gallagher, 'The socially extended mind', *Cogn. Syst. Res.*, vol. 25–26, pp. 4–12, Dec. 2013.
- [8] K. Amunts, C. Ebell, J. Muller, M. Telefont, A. Knoll, and T. Lippert, 'The Human Brain Project: Creating a European Research Infrastructure to Decode the Human Brain', *Neuron*, vol. 92, no. 3, pp. 574–581, Nov. 2016.
- [9] W. Sellars, 'Sensa or sensings: Reflections on the ontology of perception', *Philos. Stud.*, vol. 41, no. 1, pp. 83–111, Jan. 1982.
- [10] W. S. Sellars, 'Philosophy and the Scientific Image of Man', in *Science, Perception, and Reality*, R. Colodny, Ed. Humanities Press/Ridgeview, 1963, pp. 35–78.
- [11] A. Clark, 'Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head', *Mind*, vol. 118, no. 472, pp. 963–993, Oct. 2009.
- [12] K. F. MacDorman and H. Ishiguro, 'The uncanny advantage of using androids in cognitive and social science research', *Interact. Stud.*, vol. 7, no. 3, pp. 297–337, 2006.
- [13] R. Brandom, *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1998.
- [14] W. Sellars, 'III.—INFERENCE AND MEANING', *Mind*, vol. LXII, no. 247, pp. 313–338, Jul. 1953.
- [15] S. Rainey, 'Making It Problematic', in *AISB 2014 - 50th Annual Convention of the AISB 2014 50th Annual Convention of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour*, Goldsmith's University, London; United Kingdom, 2014.
- [16] J. Searle, *Making the Social World: The Structure of Human Civilization*. Oxford: OUP, 2010.
- [17] J. G. Avildsen, *The Karate Kid*. 1984.
- [18] 'Research Groups: APT - Advanced Processor Technologies (School of Computer Science - The University of Manchester)'. [Online]. Available: <http://apt.cs.manchester.ac.uk/projects/SpiNNaker/>. [Accessed: 21-Mar-2017].
- [19] 'Roboy.org | HELLO MY NAME IS ROBOY'. .
- [20] T. Lei, R. Barzilay, and T. Jaakkola, 'Rationalizing Neural Predictions', *ArXiv160604155 Cs*, Jun. 2016.