

Kinga Jęczmińska, Summary of the PhD dissertation "Consciousness beyond the Cartesian Theatre: Contemporary Anti-Cartesian Theories of Consciousness"

In the dissertation, I analyse anti-Cartesian theories of consciousness, i.e. theories that reject the model of the Cartesian theatre defined by Dennett (1991): Baars's global workspace theory, Dennett's multiple drafts model, O'Regan and Noë's sensorimotor theory and the predictive processing framework. These theories reject the assumption about the existence of some central point in the mind that each piece of information would have to go through in order to become conscious. I compare the global workspace theory, the multiple drafts model, conservative predictive coding, the radical predictive processing and the sensorimotor theory in terms of the ways in which they depart from the model of the Cartesian theatre. I also analyse the explanatory power of the anti-Cartesian theories, focusing on how they solve problems in consciousness studies without retreating to the traditional approach to qualia.

The starting point for the dissertation was the article by Degenaar and Keijzer (2009), in which the authors indicate that two theories of consciousness – the global workspace theory and the sensorimotor theory – are compatible with each other and their synthesis can explain more than each theory on its own because these theories focus on the solution of different problems of consciousness. Degenaar and Keijzer briefly outlined three possible ways of combining these two theories: the internal localisation scenario, the external localisation scenario and the no localisation scenario. In the dissertation, I analyse basic assumptions of the anti-Cartesian theories and consider all three ways of combining the global workspace theory with the sensorimotor theory. I show that it is possible to combine the global workspace theory with the sensorimotor theory within the external localisation scenario by means of the radical predictive processing developed by Clark (2015a, 2015b, 2015c, 2016).

I apply the new theory created as a hybrid mental architecture to solve basic problems of consciousness discussed in contemporary literature in philosophy and cognitive science: the concept of consciousness; integration of consciously processed information and its comprehensibility for all unconscious processors; sensations and qualia; the absolute explanatory gap and the comparative explanatory gap; differences between the first-person and third-person perspectives and the personal and subpersonal levels; the binding problem; attention and cognitive penetrability. I also apply the new theory to explain cognitive changes occurring in psychoses and autism, change blindness and inattention blindness, blindsight

and changes in perception caused by image-inverting lenses and two-coloured lenses.

In the Introduction, I briefly describe the notion of the model of the Cartesian theatre and the distinction between weak and strong anti-Cartesianism. I discuss traditional arguments that support the existence of qualia and the hard problem of consciousness. I also present counter-arguments that criticise the traditional concept of qualia. I explain the purpose of the dissertation that consists in the analysis of contemporary anti-Cartesian theories of consciousness and presentation of a combination of two of them within one of the three scenarios delineated by Degenaar and Keijzer

In chapter 1, I present Baars's theory of the architecture of the mind as the primary example of global workspace theories (Baars 1988; 1997a; 1997b; 2002; 2007, Baars *et al.* 2013). I show how this theory focuses on functions of consciousness. I present connections between the global workspace theory and contemporary research in neuroscience. I explain how the global workspace theory rejects the model of the Cartesian theatre and introduces a new modified theatre metaphor to analyse interactions between conscious and unconscious processes. I show that the theory can explain some frequently discussed issues related to consciousness studies, e.g. the stream of consciousness, Libet's experiments on free will, blindsight. I outline the limitations of the theory, which consist mainly in focusing on the absolute explanatory gap and neglecting the comparative explanatory gap.

In chapter 2, I describe the model of the mind in Dennett's multiple drafts model. I present its relations with empirical evidence concerning the functioning of the mind. I compare the multiple drafts model with Baars's global workspace theory. These two theories are similar in treating consciousness functionally, as arising from the cooperation of distributed processors that work in parallel. The multiple drafts model differs from the global workspace theory in rejecting a clear boundary between conscious and unconscious processes. I indicate that the model is highly influenced by Dennett's philosophical views, in which a significant role is played by the notions of the intentional stance, heterophenomenology and interpretivism. I discuss some limitations of the multiple drafts model, which are related to its failure to explain some crucial functions of consciousness. These issues include: the coordinative and integrative role of consciousness, some important aspects of the self and the impact of embodiment on conscious states.

In chapter 3, I present O'Regan and Noë's sensorimotor theory (O'Regan & Noë 2001a, 2001b, 2001c; Noë 2004; O'Regan 2011). I show how it departs from traditional representationalism and physicalism. I illustrate how the theory adopts basic assumptions of enactivism. I describe the strong anti-Cartesian character of the sensorimotor theory that is

related with the rejection of both the model of the Cartesian theatre and a definite division between the brain and the rest of the body. I discuss the concept of sensorimotor contingencies and their role in explaining experience on the personal level. I explain how the theory manages to bypass the hard problem of consciousness and to skilfully reformulate some traditional problems like Levine's explanatory gap by redefining qualia. The theory explains differences between various sensory modalities and differences between experiences within the same sensory modality through the notion of sensorimotor contingencies. Since the theory closes the intermodal and intramodal comparative gap but does not deal with the absolute explanatory gap, it is beneficial to combine it with the global workspace theory.

In chapter 4, I discuss some of the most important experiments and empirical data relevant for the sensorimotor theory. I indicate that the sensorimotor theory is compatible with the results from experiments on saccades, the blind spot, sensory substitution systems, change blindness and inattention blindness, image inverting lenses, two-coloured lenses and evolutionary approach. The sensorimotor theory may be described as having a higher explanatory power than alternative theories, e.g. traditional representationalism.

In chapter 5, I present predictive processing, paying a special attention to its version developed by Clark (2015a, 2015b, 2015c, 2016), which I later use to carry out the main task of my dissertation that consists in combining the global workspace theory with the sensorimotor theory. I show how the radical predictive processing explains action and perception – including interoception. This specific version of predictive coding does justice to a close relation between perception and action emphasised by enactivism that inspired the sensorimotor theory. I indicate how the radical predictive processing developed by Clark differs from more conservative versions of predictive coding. Furthermore, I present Hobson and Friston's research on dreams and development of consciousness, which is relevant for the analysis of consciousness through a metaphor of the virtual reality in predictive coding. I indicate that predictive coding can be considered an anti-Cartesian theory since it rejects the model of the Cartesian theatre. The radical predictive processing may also be an anti-Cartesian theory in a strong sense since it may also reject a clear division between the brain and the rest of the body. I explain how predictive coding can be analysed through a modified metaphor of a theatre, which is related to the model of virtual reality produced out of hypotheses about the world.

In chapter 6, I describe three scenarios distinguished by Degenaar and Keijzer (2009) as possible ways of combining the global workspace theory and the sensorimotor theory: the

internal localisation scenario, the external localisation scenario and the no localisation scenario. I present the synthesis of the global workspace theory and the sensorimotor theory by means of the radical predictive processing within the second type of synthesis among the three distinguished by Degenaar and Keijzer (2009). The combination within the external localisation scenario has a higher explanatory power than each of the theories on its own. I claim that the two other scenarios are less attractive than the selected scenario of external localisation, but I do not exclude the possibility of combining the theories within the alternative scenarios by other researchers.

Moreover, I briefly discuss some methodological issues related to the integration and unification of theories: the integration challenge in the unification of theories described by Bermúdez (2010/2011), older approaches to integration (intertheoretic reduction or analysis in terms of Marr's tri-level approach), the notion of a mental architecture and specifically a hybrid mental architecture (Bermúdez 2010/2011), and interlevel integration (Craver & Darden 2013). I also compare the global workspace theory, the sensorimotor theory and predictive coding in terms of explanatory internalism and externalism. I discuss early attempts at combining the global workspace theory with the embodied mind approach (Shanahan 2005) and conservative predictive coding with the global neuronal workspace theory (Hohwy 2013, 2015).

In chapter 7, I compare the global workspace theory, the multiple drafts model, conservative predictive coding, the radical predictive processing and the sensorimotor theory with respect to the way in which they reject the model of the Cartesian theatre. I briefly describe elements from Descartes' philosophy that are relevant for the concept of the model of the Cartesian theatre discussed by Dennett. I enumerate elements characteristic for the Cartesian mode of thinking about consciousness. I demonstrate that the analysed anti-Cartesian theories differ in their approach to the Cartesian model. I propose to analyse the theories by means of an ordinal scale on which it is possible to assess their different modes of departure from the model of the Cartesian theatre. These theories modify the original Cartesian theatre metaphor and introduce their own modified theatre metaphors. I show how the predictive coding framework can constitute the link between two kinds of models: the conservative model adopted by the global workspace theory and partially by the multiple drafts model and the radical model supported by the sensorimotor theory.

In chapter 8, I explain how the unified account can approach problems that were previously discussed from independent perspectives of analysed theories. The complex

account of consciousness, which comprises elements of the global workspace theory and the sensorimotor theory unified by the radical predictive processing within the external localisation scenario, is capable of providing a fuller explanation of problems usually posed with reference to consciousness: the concept of consciousness; integration of consciously processed information and its comprehensibility for all unconscious processors; sensations and qualia; the absolute explanatory gap and the comparative explanatory gap; the distinction between the first-person perspective and the third-person perspective; the problem of sensory binding; cognitive penetrability; changes in psychoses and autism; attention; experiments on change blindness and inattentional blindness; blindsight; changes in perception caused by image inverting lenses and two-coloured lenses. The combination of the global workspace theory and the sensorimotor theory by means of the radical predictive processing results in a comprehensive description of these exemplary issues.

The presented hybrid mental architecture comprises elements of several contemporary approaches to consciousness. Although each approach has its strengths and can be used independently to focus on particular problems, on a more general conceptual level it may be beneficial to study consciousness in a more comprehensive manner.