Original intentionality and linguistic meaning: an historical outline

Hermann Moisl
Newcastle University, UK
hermann.moisl49@gmail.com

Introduction

A long-standing problem in linguistics and cognitive science more generally is how natural language expressions come to possess, and how artificially intelligent systems can be endowed with, intrinsic meaning. There is a long tradition in Western thought whereby (i) the meanings of linguistic expressions are their significations of mental concepts, (ii) mental concepts include representations of the mind-external environment causally generated by the cognitive agent's interaction with that environment, and (iii) these representations are intrinsically meaningful by virtue of their connection of mind and world, thereby providing the basis for intrinsic linguistic meaning.The present paper outlines this tradition with the aim of providing the intellectual context in which cognitive models of intrinsic linguistic meaning can be constructed.

The discussion is in two main parts: the first motivates engagement with this topic, and the second presents the historical outline.

1. **Motivation**

In the second half of the twentieth century, cognitive science in general and linguistics in particular were dominated by the Computational Theory of Mind (CTM), whereby the mind is seen as a Turing Machine whose program is cognition (Rescorla 2020). Its dominance in recent decades has been challenged by neural (Churchland 2012) and dynamical systems (Ward 2002**)** approaches to cognitive theorizing, but with respect to meaning in particular the most fundamental challenge has come from a philosophical thought-experiment formulated by the philosopher John Searle in 1980 and subsequently developed by him (Searle 1984, 1989, 1990, 1999, 2002, 2010). Searle's initial aim was to counter the claim by 'strong' artificial intelligence that it is possible to construct machines with human-level intelligence by basing their design on CTM, but he subsequently extended the implications of the experiment to the philosophical foundations of CTM itself, arguing that it cannot in principle offer a complete theory of cognition. This attack on *de facto* standard cognitive science and its application in AI precipitated an extensive controversy which continues to the present day, but to date there is no consensus either on the validity of his position or, assuming that it is valid, whether or how AI and CTM might adapt to it (Cole 2020).

Searle's argument (1980) is based on two propositions:

> (1) '*Intentionality in human beings (and animals) is a product of causal features of the brain'.*
> (2) '*Instantiating a computer program is never by itself a sufficient condition of intentionality'.*

He takes (1) as '*an empirical fact about the actual causal relations between mental processes and brains. It says simply that certain brain processes are sufficient for intentionality',* and (2) as something to be established by argument, which he undertakes. The conclusions are that 't*he explanation of how the brain produces intentionality cannot be that it does it by instantiating a computer program',* which is '*a strict logical consequence'* of (1) and (2), and that 'a*ny mechanism capable of producing intentionality must have causal powers equal to those of the brain. This is meant to be a trivial consequence of 1. Any attempt literally to create intentionality artificially (strong AI) could not succeed just by designing programs but would have to duplicate the causal powers of the human brain'.*

These conclusions assume the validity of (2), the argument for which is based on the Chinese Room thought experiment; the version of the experiment given here is that in (Searle 1999). There is a closed room containing Searle and a list of rules in English for manipulating Chinese orthographic symbols. Chinese speakers outside the room put sequences of these symbols into the room and, using the rules available to him, Searle assembles and outputs sequences of Chinese symbols in response. The people outside interpret the input sequences as sentences in Chinese whose meaning they understand and the output sequences as reasonable responses to them, and on the basis of the room's conceptually coherent input-output behaviour conclude that it understands Chinese. Searle himself, however, knows that the room does not understand Chinese because he, the interpreter and constructor of the sequences, does not understand Chinese, but is only following instructions without knowing what the input and output strings mean.

Key to interpretation of the experiment is the philosophical concept of intentionality (Jacob 2019). Etymologically it is related to Latin *intendere*, 'to point at, to direct', and was used in medieval European philosophy to refer to the mind's ability to direct its attention to specific mental concepts as well as to things and states of affairs in the mind-external world. In present-day philosophy of mind *intentionality* is used to denote the 'aboutness' of mental states, '*the power of minds and mental states to be about, to represent, or to stand for, things, properties and states of affairs*' (Jacob 2019). Searle distinguishes two types of intentionality, original and derived, where the locus of original intentionality is the human head, and derived intentionality is that which we attribute to physical mechanisms which we have good reason to believe do not have original intentionality, such as thermostats, whose operation is routinely interpreted by humans as 'wanting' to maintain an even temperature.

The Chinese Room is, of course, a computer. Searle is the CPU, the list of English instructions is a program, and the input / output sequences are symbol strings; by concluding that the room understands Chinese, its observers have confirmed the Turing Test (Oppy & Dowe 2016), which says that any device which can by its observable behaviour convince human observers that it has human-level cognition, that is, intentionality, must be considered to possess it. Searle knows, however, that the room's intentionality is derived, not original, the implication being that physical computer implementations of CTM models in AI systems, like thermostats, can only ever have derived intentionality. The intentionality of the symbols and symbol structures manipulated by the algorithm of a CTM model are in the heads and only in the heads of their human designers. When physically instantiated, for example by compilation of a program onto a physical computer, this intentionality is lost: the symbols and symbol structures cease to be symbolic and become physical tokens which drive the physical causal dynamics of the machine, but intentionality is not a factor in that causal structure. If well designed, the behaviour of the machine with respect to whatever aspect of cognition the CTM algorithm was intended to model can be interpreted as intentional, just as the behaviour of the Chinese room can be so interpreted, but once again the semantics is derived because the only locus of intentionality is in the heads of observers. Put simply, a physical computer does not in any sense understand what it is doing any more than a vending machine does; it simply pushes physical tokens around, and humans interpret that activity as intentional.

Given the lack or consensus on Searle's position or prospects of one, an alternative approach to *a priori* philosophical discussion of it is empirical: assume that Searle is right about the original / derived intentionality distinction and about the view that a physical system can only have original intentionality if intentionality is a causal factor in its operation, and construct models based on those assumptions to see if any useful scientific insights ensue. The present discussion takes a first step in that direction by reviewing the history underlying these assumptions.

2. **Historical outline**

The tradition referred to in the Introduction has a two and a half millennium-long history that begins with the pre-Socratic philosophers of ancient Greece and finds its current expression in the various academic disciplines that comprise cognitive science. A thorough review of this body of thought would require engagement with truly vast associated literatures, and would amount to a history of a large part of Western thought. Even assuming the requisite authorial competence, the sheer size of such a review would clearly rule it out here. What follows is therefore a sketch, the aim being to provide a conceptual context for model construction. To forestall a bibliographical deluge, the policy is to cite recent work containing further references.

2.1 **Antiquity (c.600 BC - c.500 AD)**

Philosophical interest in the nature of linguistic meaning is documented in the Ancient Greek cultural world from the 6th century BC onwards (Allan 2009). The framework for discussion of was predominantly psychological in that language was analyzed as an aspect of mind. The most important figures were Aristotle (384–322 BCE) , Augustine (354–430 CE), and Boethius (c.480–524 CE).

Aristotle's views on meaning in language have been hugely influential in Western philosophy (Cohen 12016; Modrak 2001; Shields 2012, 2014, 2015, 2016). They can only be fiully understood in the context of complex ontology, epistemology, and philosophy of mind developed across the corpus of his work, but the essence for present purposes is this. The mind has two main components: the perceptual faculty which gives it access to the external world via the senses, and the noetic faculty which synthesizes and interprets the output of the perceptual one. The perceptual faculty is innate and includes both sensation and imagination, where the former, as its name suggests, deals with the senses and their role in providing input from the mind-external world, and the latter with synthesis of sensory input. His epistemology is based on the assumption that the structure of independently-existing mind-external physical reality is accessible to the perceptual faculty, and his account of perception is framed in terms of his concept of hylomorphism, 'matter-formism', whereby a natural substance comprises matter and form such that the form is what differentiates the matter of any one thing from any other. The form but not the matter of a mind-external object enters the perceptual faculty via the senses, where it is represented as an image; Aristotle's analogy is a signet ring pressed into hot wax, which leaves the form of the design on the the ring's face in the wax, but not the gold from which the ring is made. The perceptual faculty not only represents mind-external objects in this way, but also abstracts over formally-similar representations using the 'active intellect' innate in all humans, so that, for some collection of physically similar forms in external reality, it generates a representation of the essential form, or universal, underlying the variant physical instantiations; such a representation is an *eikon*, an image that represents not particulars of an object in reality, but an abstraction of a class of similar objects. And, finally, Aristotle is clear that words signify the mental representations generated by the perceptual faculty and that this signification is conventional in the sense that it is based on societal consensus.

Augustine (Tornau 2019; Meier-Oeser 2011; Klima 2017) proposed a theory of signification in which a sign is "*something that shows itself to the senses and something other than itself to the mind*", and can be one of two kinds, conventional and natural, the latter of which '*apart from any intention or desire of using them as signs, do yet lead to the knowledge of something else*', that is, things which by their nature signify without requiring conventional agreement, such as footprints signifying the passing of an animal (quotations from Meier-Oeser 2011). With reference to language, spoken words in any given language are conventional signs of the speaker's mental concepts which the speaker uses to convey those concepts to other minds. Concepts are generated by a combination of sensory experience of the mind-external world and divine illumination, and, following Aristotle, are natural signs of things and events in the world - there is a causal, nonconventional connection between states of the world and how they are represented in the mind, and consequently a resemblance between the structure of mind-external reality and mind-internal conceptual structure. Words in the various spoken languages conventionally signify their counterparts in the mental language which, also following Aristotle, is common to all humans in the sense that humans all experience the world in closely similar ways because we share the same kinds of mind and sensory apparatus as part of human nature.

Boethius (Klima 2017; Marenbon 2016; Meier-Oeser 2011) was the main conduit of Aristotelian philosophy to medieval Europe. His main personal contribution lies in his attempt to reconcile the empirical Aristotelian view of universals with that of Augustine, whose view was strongly influenced by the neoplatonic tradition of abstract Forms or Ideas existing independently of human mentality. The problem had to do with the semantics of so-called common names. Proper names and indexicals can refer directly to particular entities in the physical world (Ashworth 2015), but what are the referents of common terms that refer to things which have no physical existence in the world - not only words for obvious abstractions like 'truth', but also ones which pervade natural language and denote types of real-world object such as 'man'? Boethius took an Aristotelian empiricist view, whereby universals are mental concepts derived by abstraction from sensory experience, and cites with approval the view of Alexander of Aphrodisias (*floruit* c.200 CE) that spoken words do not signify things in the world but our mental concepts of such things, and that universals have no independent existence in the mind-external world but are mental abstractions of sensory experience generated by the Aristotelian active intellect: spoken words conventionally signify concepts, and concepts nonconventionally signify states of the world.

2.2 **The European Middle Ages (c.500 - c.1500 AD)**

There is nothing of importance for present purposes in the period from Boethius until after 1000 AD.This changed thereafter for two main reasons. One was the growth in academic activity in newly-established European universities. The other was increasing availability of the works of Aristotle via contact with the Islamic world (Shields 2015; Spade 2016). In combination, these developments revolutionized philosophical discourse and led to a substantial amount of work on the theory of mind and language (Allan 2009; King 2007; Meier-Oeser 2011; Spade 2016).

Anselm of Canterbury (1033-1109 CE) combined Aristotelian and Augustinian views on concepts and linguistic meaning in mental words, which have natural signification and are '*resemblances and images of things*' (Meier-Oeser 2011). It was, however, his younger contemporary Peter Abelard (1079–1142 CE) who pioneered subsequent developments in medieval philosophy of mind and language (Allan 2009; King 2007; King & Arlig 2010). His ideas were based on a sparse ontology that was strongly Aristotelian and therefore physical monist and reductive, whereby the only real things in the mind-external world are physical particulars. Following Aristotle, thought is based on concepts which are derived from sensory experience of the world and which represent that experience in the mind. These concepts are nonconventionally related to the world in that there is a causal connection between states of the world and how they are represented in the mind, and they are common to all humans in that, relative to some given state, all humans represent that state in similar ways because we share the same kind of mind and sensory apparatus as part of human nature. Experientially-derived concepts are the basis for the mind's creation of abstract concepts which have no existence in mind-external reality. In arguing for this view of concepts Abelard had to resolve a problem transmitted to him by Aristotle via Boethius and deriving from his ontology: given that, to have meaning, words must refer ultimately to particulars in the world, what is the reference of a so-called common name, that is, of universal words like 'man'? There are particular men in the world, but no universal man. Abelard's solution was twofold. Firstly, he proposed that common names do not signify universals but rather are universals themselves - universals are words, not things that words signify (King 2007; King & Arlig 2010; Klima 2017; Rodriguez-Pereyra 2014). And secondly, anticipating Frege, he proposed a binary decomposition of Augustinian signification into sense (*significatio*) and reference (*nominatio*). The *nominatio* of a proper name like 'Aristotle' is a reference to a particular man in the world, and of a common name like 'man' is the particulars to which the word applies by distributive reference, or, in Aristotle's formulation, the word is 'predicable of many'; the *significatio* of a proper name is a sense-derived concept, and of a common name an abstract concept. The meaning of a word is thereby a combination of its *nominatio* and its *significatio.* Underlying all this, finally, is the traditional principle that both the sense and the reference of a word in spoken language are conventional.

Abelard initiated an extensive discussion of semiotic and specifically linguistic semantic topics based on a complex interplay of Aristotelian and Augustinian thought which lasted until the end of the Middle Ages and beyond (Allan 2009; Meier-Oeser 2011; Shields 2015; Spade 2007, 2016). There are too many important figures in that discussion to deal with individually, so topics of particular relevance to present purposes are outlined instead.

- Universals

  The ontological status of universals fell into three broad categories derived from Antiquity: that they are *universalia ante rem*, 'before the thing', that is, as independently existing Platonic objects, or *universalia in re*, 'in the thing', as somehow implicit in particulars, or *universalia post rem*, 'after the thing', as concepts without existence independent of the mind's abstraction over particulars (Klima 2017). The main disagreement was *in re* versus *post rem,* or in present-day terms between realists and nominalists respectively; the *universalia post rem* position had become dominant by the 14th century with the resurgence of the nominalism pioneered by Abelard and associated primarily with William of Ockham (c. 1287–1347; Meier-Oeser 2011; Spade 2016), whereby the only universals are words that signify concepts, and concepts are mental representations that signify particulars in the physical world.

- Signification
  Signification was fundamental in Abelard's philosophy of language and remained so in medieval semiotics and linguistic semantics (Meier-Oeser 2011; Spade 2007), as did the distinction, derived from Aristotle and Augustine, of natural and conventional signs, where natural signs are causally generated in the mind by perception of states of the world. There is a three-level hierarchy of signification for words (Spade 2007): a written word signifies the corresponding spoken word, a spoken word signifies the corresponding mental concept, and a mental concept signifies a mind-external object or state of affairs in the natural world.

- Mental representation
  Aistotle's view of mental representation was that representation of the world external to the cognitive agent was essentially pictorial. It was not until the 13th century that the idea of mental imagery was questioned by, for example, Aquinas (1225-1274), who used the metaphor of a blueprint for a building to argue for a more abstract notion of mental encoding rather than direct resemblance (Lagerlund 2017; Spade & Panaccio 2019).

2.3 **The modern era (c.1500 - present)**

The medieval 'scholastic' tradition was gradually supplanted by the rise of empirical science in the course of the 16th, 17th, and 18th centuries (Allan 2009; Gendler Szabo 1998; Klima 2017; Meier-Oeser 2011), in which thinking on language was strongly influenced by the new scientific spirit. An early example was the *Port Royal Grammar* of 1660 (Buroker 2014; Gendler Szabo 1998), whose fundamentals were that (i) thought is prior to language, (ii) proper names signify ideas of particulars, and common names signify ideas of universals, and (iii) the relationship between linguistic expressions and ideas is conventional, but between ideas and what they represent in the mind-external world is not. An approximate contemporary was John Locke (1632 - 1704; Allan 2009; Gendler Szabo 1998; Uzgalis 2018), who famously maintained that the mind is a blank slate at birth on which experience writes knowledge of the mind-external world. Experience is of two kinds, sensation and reflection, where the first is information about the world provided by the senses, and reflection is a mental process that builds new, more complex ideas from existing ones, both sensory and other products of reflection; the mental process is a set of innate 'faculties' whereby ideas are structured. Proper names signify ideas based on perception, and common names signify universal ideas constructed by the mind's 'faculties'.

The period from the end of the eighteenth century onward has seen rapid development of empirically based science and technology, and developments in the study of language have mirrored that empirical orientation. John Stuart Mill, for example (1806-73; Macleod 2016), regarded the mind as part of the causal order of physical nature amenable to study by the empirical methods of the natural sciences rather than as a metaphysical entity. Understanding of the world comes solely from inductive inference from sensory experience: sense impressions frequently experienced simultaneously or in immediate succession are associated to become ideas and causal relations among ideas respectively. This process is iterative, so that as experience of the world grows over time, the ideational and causal structure of the mind increasingly comes to resemble the corresponding structure inherent in mind-external reality. From these fundamentals the mind constructs complex ideas and relations all of which, however abstract, are ultimately grounded in experience; language expresses these complex ideas and relations. Another example is Charles Peirce (1839-1914; Atkin 2010), who developed a theory of signification, representation, reference, and meaning which became the foundation of the present-day discipline of semiotics. The fundamental idea is that a sign has three components: a signifier, that which is signified, and an interpreter who connects the two. Signs can be of three kinds: (i) an icon is a sign whose signifier bears a resemblance to what it signifies, like a painting of a person, (ii) an index is a sign whose signifier is causally connected to what is signified, like smoke signifying fire, and (iii) a symbol is a signifier whose connection with what it signifies is conventional.

The 19th century saw the first attempts to understand the relationship of mind and brain and of human sensory systems from a scientific rather than purely philosophical point of view, as seen in the work of, for example, E. H. Weber (1795-1878), Gustav Fechner (1801-1887), Hermann von Helmholtz (1821–1894) , Ernst Mach (1838–1916), and Wilhelm Wundt (1832-1920). Of these, von Helmholtz and Mach are of particular relevance to present concerns on account of their views on the genesis of perceptually-based mental representations.

- Von Helmholtz (Patton 2018) was a physicist with a strong interest in perception as the basis for scientific epistemology, and was a pioneer of the view that epistemology is crucially dependent on the dynamic interaction of body and environment. The essence of his view was that percepts were 'signs' which symbolize sensory stimuli, but that, unlike in earlier so-called picture theories of perception, there is no necessary direct resemblance between the form of the stimuli and their perceptual representation. Physical stimuli occur in temporal sequence, and the mind makes 'unconscious inferences' from the sequence of percepts to construct a temporal ordering and thereby a coherent mental representation of experience whose structure reflects regularities in the physical stimuli on which it is ultimately based. Perception of space and sequence is determined by constraints that the structure of the perceiver's body places on interaction with the physical environment: *'perceptual space is a mental generalization of our orientation with respect to objects in space*' (Patton 2018).
- Ernst Mach (Pojman 2019) was a physicist and mathematician with a conviction of the need to take the body into account when studying perception. On his account, sensory systems are dynamical systems tending to equilibrium with the environment by a continual process of adaptation to physical stimuli over time; what the senses thereby make available to the brain are not direct representations of reality but generalizations over experiential history, with the result that there is no necessary isomorphism between percept and reality, or, in other words, human access to knowledge of mind-external reality is indirect. This tendency to equilbrium, moreover, is teleological in the Darwinian sense of fitness for an environment leading to optimization of survival chances. Mach's empiricism was not of the Lockean 'blank slate' variety, but had the apparently-paradoxical effect of allowing him to admit a priori structure to his account of perception in that, like evolutionary processes in nature more generally, the human sensory apparatus had evolved and continues to evolve towards equilibrium with a structured mind-external reality.

Since the mid-20th century the study of intentionality and of linguistic meaning specifically have been addessed by a range of disciplines, chief among them linguistics and philosophy of language, philosophy of mind, cognitive psychology, and cognitive neuroscience.

2.3.1 **Linguistics**

Linguistics in the second half of the 20th century has, at least in the USA and the UK, been dominated by the generative framework initiated and thereafter guided by Chomsky (Newmeyer 1995). It has been and continues to be strongly orientated towards study of syntax; semantics consists essentially of application of ideas from the philosophy of language (Lycan 2018), which has in turn developed the tradition of truth-conditional semantics initated by Frege. This approach regards language, and with it meaning in language, as an abstract object to be understood using mathematical logic independently of its instantiation in human minds; natural language is regarded as one instance of a universe of possible languages to the understanding of which general semantic principles such as reference and compositionality can be applied. Because of its apsychological orientation, it is of peripheral interest for present concerns. There are exceptions to the dominance of truth conditional semantics in linguistics, however.

- Generative semantics (McCawley 1976, 1995) was an alternative to the standard generative linguistics view of the relationship between syntax and semantics developed in the late 1960s to the mid-1970s. It proposed that, contrary to the standard generativist position, deep structures were semantic rather than syntactic, which transformational rules converted into surface strings and associated syntactic structures. The semantic subcomponent thereby made meaning primary in sentence generation.
- Conceptual Semantics (Jackendoff 2002, 2015) emphasizes the relationship between language and other aspects of cognition, which among other things involves a demotion of syntax from the centrality accorded it in generative linguistics to a subsidiary role. In Jackendoff's words, 't*he central hypothesis of Conceptual Semantics is that there is a level of mental representation, Conceptual Structure, which instantiates sentence meanings and serves as the formal basis for inference and for connection with world knowledge and perception'.* Conceptual structure interfaces with two subsystems: a perception / action subsystem that generates representations of the mind-external world, and a linguistic input / output one that generates phonological and syntactic representations of acoustic input and of acoustic output that a listener with the same cognitive system can interpret as meaningful. Reference is a relation between a mental representation in the linguistic input / output subsystem and some part of conceptual structure: linguistic expressions do not refer directly to things in the world but indirectly via mental representations in conceptual structure wholly or partly generated by interaction of conceptual structure with the perception / action subsystem.
- Cognitive linguistics (Croft & Cruse 2004; Geeraerts & Cuyckens 2010) grew out of the generative semantics of the 1970s, and is now an umbrella term for a wide range of approaches. It resembles Conceptual Semantics in locating meaning in general cognition rather than in the linguistic system specifically, and in stressing the mapping from cognitive meaning to linguistic structure as the primary focus of linguistic theory. The currently dominant variety of this approach to language is Cognitive Grammar (CG) (Croft & Cruse 2004; Langacker 2008; Talmy 2000). Geeraerts & Cuykens (2007) identify three fundamental tenets of CG: (i) meaning is primary in linguistic analysis, (ii) meaning draws selectively on the totality of the individual speaker's world knowledge, and (iii) linguistic utterances communicate not objective truths about the world, but rather the speaker's perspective on it.

2.3.2 **Philosophy of mind / cognitive psychology**

Study of the psychological mechanisms of intentionality has largely been the preserve of philosophy of mind and cognitive psychology. Topics of particular relevance here are perception, representation, and cognitive architecture.

Perception

Perception is the cognitive process by which humans acquire knowledge of the mind-external world (Crane & French 2015; Lyons 2016; Siegel 2016). A central problem has been skepticism (Comesana 2019; McKinsey 2018), which is the view that we cannot have knowledge, that is, true justified belief, of anything about the mind-external world on account of the unreliability of perception. This remains a problem, but, setting it aside, proposals for how we acquire knowledge of the world have been on a continuum between the rationalistic view that such knowledge is innate, where the knowledge intrinsic to the mind determines our perception of the world and thereby creates our reality, and the empiricist view that knowledge of the world is derived solely from sensory experience of a mind-independent physical reality (Adams & Aizawa 2017; Margolis & Laurence 2019; Markie 2017; Pitt 2012; Samet 2019; Samet & Zaitchik 2017). Few philosophers and psychologists have been pure innatists or pure empiricists (Griffiths 2009; Samet & Zaitchik 2017). Instead, most have taken intermediate positions: on the one hand, innatists have found it necessary to accommodate the self-evident fact that we use our senses to perceive the world, and have proposed sensory perception as a trigger which brings the innate knowledge to consciousness, and, on the other, empiricists have found it necessary to posit an innate mental capacity for induction of knowledge from sensory experience. Most recent and contemporary cognitive scientists subscribe to some combination of innate and environmentally-derived empirical factors as the bases for the acquisition of world knowledge (for example Barsalou 2016a; Griffiths 2009; Lau & Deutsch 2014; Margolis & Laurence 2019; Petersson & Hagoort 2012; Samet & Zaitchik 2017; Shea 2018; Wilson & Foglia 2015).

The nature of this interaction has in recent decades been developed by a variety of research programmes including naturalistic (Rysiew 2016), evolutionary (Bradie & Harms 2016), and teleological epistemology (Neander 2012, 2017; Shea 2018), evolutionary psychology (Buss 2007; Downes 2014), and embodied / situated / grounded cognition (Barsalou 2008, 2010, 2016a; Clark 2008; Lakoff & Johnson 1999; Matheson & Barsalou 2018; Wilson & Foglia 2015; Yeh & Barsalou 2006;). Though differing in focus, the core ideas of these programmes overlap substantially. The guiding principle is that philosophical discussion of knowledge acquisition should work closely with relevant work in the natural sciences rather than relying exclusively on the traditional a priori philosophical method. In particular, they have adopted the adaptationist position in the biological sciences (Orzack & Forber 2010), whereby natural selection is an important and probably the primary causal factor in the adaptation of living organisms to fit their specific environmental niches: human cognition is explained in terms of interaction between evolutionarily-generated cognitive mechanisms, physical environment, and individual learning of culturally-transmitted behaviour and knowledge (Barsalou 2016a; Shea 2018). The traditional opposition of innate versus empirical as factors in the acquisition of world knowledge is replaced by a dynamical process which unifies them whereby, by adaptation over evolutionary time, the physical and cultural environment shapes cognitive structure, which in turn interprets and acts on the environment.

Representation
Like Descartes, my individual consciousness tells me that I have a mind, that my mind is constituted by thoughts, and that most of these thoughts are about my relationship with the mind-external world. Intuitively, and in the history of philosophy of mind specifically, various terms have been used to designate the components of thought, such as 'phantasms', 'ideas', 'impressions', 'notions', and 'concepts'. In everyday usage the definitions of these terms and their interrelationships are vague, and in the academic literature writers typically use them in different ways at different times and places, the result of which is terminological confusion. The present discussion follows Margolis & Laurence (2019) in standardizing on 'concepts' as *'the building blocks of thoughts*'.

The historical view of concepts as pictures in the mind is long-obsolete, and the dominant view in present-day cognitive science, articulated in the Representational Theory of Mind (RTM; Pitt 2012), is that concepts are mental representations. Because of their central importance in cognitive science there is an extensive literature on the nature of these representations, reviewed for example in (Adams & Aizawa 2019; Carey 2009; Margolis & Laurence 2015, 2019; Pinker 2007; Pitt 2012), which precludes a comprehensive summary here. Instead, what follows identifies some mainstream features and then goes one to describe one particular class of theories: the causal view of mental representation.

In semiotic terms, the move from mental pictures to representations is from icon to symbol. Representations are symbols, and as such there is no necessary formal resemblance between a representation and what it represents. Representations are physically individuated tokens in the heads of cognitive agents, and each token has a semantic interpretation, where the referent or 'content' is a state of the world or another head-internal representation. In this way, representations connect the agent with the external world, and this connection is the foundation on which intentionality is constructed in the agent's mind. Representations are typically regarded as structured mental entities of arbitrary complexity, compositionally constructed from primitives which are either innate or empirically inferred via perception or some combination of the two. An influential view, proposed at various times over the millennia and reintroduced most recently by Fodor (1975), is that the mind's representational system is a language - a language of thought - in which primitive representations are analogous to words in natural language, whose forms are independent of the forms of what they represent and which can be compositionally assembled into complex expressions.

Of particular relevance to the main theme of the present discussion are causal theories of mental content, which '*attempt to explain how thoughts can be about things...These theories begin with the idea that there are mental representations and that thoughts are meaningful in virtue of a causal connection between a mental representation and some part of the world that is represented. In other words, the point of departure for these theories is that thoughts of dogs are about dogs because dogs cause the mental representations of dogs*' ; the general principle is that any given symbol 'X' means X because 'X's are caused by Xs (Adams & Aizawa 2017). Proposed causal factors include cognitive development via the individual cognitive agent's interaction with a structured physical environment under so-called normal conditions, and the development of cognitive functions by natural selection in the human genome consequent on such experience. Stampe (1977) suggested that this causal connectedness implies a homomorphism between mental representational and mind-external physical environmental structure, and subsequent work (or example Adams & Aizawa 2017; Piccinini 2018; Piccinini & Scarantino 2011; Rupert 2008; Shagrir 2012; Thomson & Piccinini 2018) has extended this idea by distinguishing natural and non-natural information and arguing that it enables non-derived or, in Searle's terms, original intentionality in physical systems, including the brain.

Cognitive architecture
The dominant view since the second half of the twentieth century has been that the mind is a computer. This is an ontological claim, not just a metaphor: the mind actually is a computer, where 'computer' is understood as a Turing Machine, a mechanism for algorithmic string transformation. The essence of this view of mental architecture, known as the Computational Theory of Mind (Boone & Piccinini 2016; Piccinini 2007, 2016; Rescorla 2020), is this:

- The brain implements a Turing computational architecture.
- The mind is an algorithm running on that architecture.
- Concepts are data structures of arbitrary degrees of complexity manipulated by the mental algorithm.
- The primitive data types from which the conceptual data structures are built are representations of mind-external reality.

In its contemporary default form, CTM is closely associated with the Representational Theory of Mind outlined above. When combined with RTM, CTM sees the mind is a computer that transforms strings of primitive representations: the mental program interprets such strings as having compositional syntactic structure of arbitrary complexity, and the nature of the transformation in any particular case is sensitive to this structure. Fodor famously referred to CTM as 'the only game in town' (ref), and it's is easy to see why. CTM is a physicalist theory, and hence one compatible with mainstream science, that explains how a physical system such as a computer or, in humans, the brain, can generate semantically coherent cognitive behaviour from the causal structure of a physical symbol system; for the classic defense of CTM see (Fodor & Pylyshyn 1988).

Since the mid-1980s the dominance of CTM has been challenged by two alternative approaches to modelling the mind: cognitive neuroscience and dynamical systems theory. The first of these is discussed in the next subsection. The second is based on the observation that nonlinear response to input and output latency in individual neurons together with pervasive feedback connectivity among biological neurons and neural assemblies make the brain a physical nonlinear dynamical system capable of behaviours characteristic of such systems, ranging from fixed-point through periodic and fractal to chaotic. The mathematical theory of dynamical and complex systems has consequently been proposed as an alternative or at least an adjunct to CTM (Churchland 2012; Metzger 2017; Port & van Gelder 1995; Ward 2001).

2.3.3 **Cognitive neuroscience**
The rapid development of cognitive neuroscience and neural modelling in recent decades has been providing ever more detailed insight into the mechanisms of the brain's implementation of traditional ideas about natural original meaning. Cognitive neuroscience attempts to identify correlations and, ideally, causal relationships between the various aspects of cognition and the anatomy and dynamics of the brain in the cognitive agent's interaction with the environment (Boone & Piccinini 2016; Gazzaniga et al 2019; Piccinini & Bahar 2013; Piccinini & Shagrir 2014). Empirical results come, on the one hand, from anatomical study and from observation of brain activity correlated with cognitive activity using the various monitoring technologies, and on the other from neural modelling (Buckner & Garson 2019; Schmidhuber 2015) using simplified artificial neural network (ANN) models of neural structure and dynamics to study the behaviour of different architectures relative to cognitive function. These approaches are complementary and often combined.

Because CTM explains cognition in terms of computation over representations, cognitive neuroscience has been particularly interested in implementation of representations in the brain (Barsalou 2016b; Boone & Piccinini 2016; Wilson-Mendenhall et al 2013). An emerging view is that representations are dynamic neural activation patterns distributed over disparate areas of the brain which are proximately or ultimately based on structures and processes in the brain's sensimotor areas generated by interaction with the environment (Barsalou 2017; Conway & Pisoni 2008; Harris et al 2001; Prinz 2002), and that a hierarchy of association areas or 'convergence zones' integrates sensimotor activations from the various modalities and other association areas to generate increasingly abstract representations (Anderson 2010; Barsalou 2016a,b, 2017; Binder 2016; Binder et al 2005, 2009, 2016; Wilson-Mendenhall et al 2013). There is, moreover, evidence for homomorphism between the structural relations among neural states and the states of the external environment they represent, which arises from a causal interaction between cognitive agent and environment and which constitutes a model of the environment (Matheson & Barsalou 2018; Neander 2017; Piccinini & Bahar 2013; Piccinini 2009, 2015, 2018). Such association areas provide a plausible implementation mechanism for the age-old problem of how sensory input is integrated in the mind so as to generate abstract concepts or, in medieval terms, universals. Barsalou & Dutriaux (2018), for example, propose the Situated Conceptualization Framework to explain how such itegration works in the head. Representations grounded in temporally co-occurring sensory and enactive experiences of the world are associated in the convergence zones, thereby generating concepts which can in turn be associated to generate increasingly abstract concepts; the conceptual structures so generated are stored in memory and interact with the cognitive agent's subsequent experience of the world, thereby dynamically constructing the agent's conceptualization of the world throughout life.

Neurolinguistics, or alternatively biolinguistics (Boeckx & Grohmann 2013; Bookheimer 2002; Friederici 2011; Kemmerer 2015), is a subfield of cognitive neuroscience which attempts to relate brain structures and processes to the comprehension, production, and acquisition of natural language. Lenneberg's influential *Biological Foundations of Language* (1967) identified biological issues relevant to the study of language that complemented Chomsky's abstract computational approach. Since then the development of electrophysiological and brain imaging technologies has generated a huge volume of empirical results which have allowed the brain to be mapped with respect to correlations between language tasks and neural activity. Such studies have established that the traditional view, which locates language exclusively in Broca's and Wernicke's areas, is only part of the answer. In common with the general multifunctional organization of brain regions mentioned above, the language network also includes the left-lateralized areas of the inferior frontal and the temporal lobes in which they are embedded as well as more distant parts of the frontal, parietal, and occipital lobes (Bookheimer 2002; Friederici 2011; Kemmerer 2015); Broca's area in particular is involved in a variety of visual and motor functions (Anderson 2010). It has also been possible to distinguish the areas specific to semantic processing within the more general language areas, and also from areas specific to sensory input modality processing (Binder et al 2009).

With respect to word meaning (Binder et al 2009; Kemmerer 2015, chs. 10-12), the closely related Grounded Cognition and Hub-and-Spoke models are gaining increasing acceptance. In the first of these the referents of linguistic expressions are mental representations of mind-external reality, and mental representations are based ultimately on perceptions of that reality as mediated by the various motor and perceptual modalities; the neural implementation of representations is held to be based on the physical activations of these modalities in response to external stimulation and motor interaction with the environment. The Hub-and-Spoke model goes on to claim that the activations of modality-specific regions are physically connected to and integrated in representations located in the temporal cortices, and that these synthetic representations are the referents of words, that is, word meanings; the representations are the hub and the modality-specific representations are the spokes; the hub integrates cortically distributed sensory and motor features of the mental representations to which words refer.

Influential examples of neural cognitive modelling are Ryder's SINBAD (Ryder 2004), which combines ANN architecture with teleological functionality to learn representations of environmental regularities via sensory input the structure of which becomes homomorphic with the environment over time, and State Space Semantics, which Paul and Patricia Churchland have developed over several decades and which is comprehensively stated most recently in Churchland (2012); for critiques see (Fodor & Lepore 1996, 1999) and for defenses (Laakso & Cottrell 2000; Shea 2018). With the explicit aim of supplanting CTM as the preferred explanatory paradigm for cognition, the Churchlands propose a dynamical systems account of how what is known of brain physiology and temporal processing generates cognition, where that system comprises a collection of interacting artificial neural networks. The essential features are as follows.

- Assuming a structured evironment and beginning with the neonatal brain, a stimulus from the environment to a sensory modality generates a pattern of neural activation in the brain. With repeated presentation of any given stimulus A, dendritic and synaptic growth maps A to activation in a specific brain location. Cognitively this is learning; from a dynamical systems viewpoint the location is an attractor in the physical brain activation space. Via the same mechanism, stimuli B, C, D... are mapped to brain locations whose distance from one another is homomorphic with the similarity structure of the stimuli. As learning accumulates for a large number of stimuli, an activation structure emerges in which similar stimuli generate activation in closely adjacent brain locations, and dissimilar stimuli activate relatively more distant locations. Each resulting concentration of activation locations is, in dynamical systems terms, a basin of attraction to which any future stimuli of a similar type are attracted. Over time, for each sensory modality, there emerges a structure which is a map of basins of attraction. Such maps are modality-specific representations of the external environment.

- Connections exist not only from sensory input modalities to brain maps, but also between maps and to association areas, the last-mentioned of which are activation areas which learn attractors from the co-activation of the sensory areas to which they are connected. These association areas implement cognitive concepts.

- The configuration of numerous interconnected sensory and association maps is formed in early life, and their representation of the mind-external environment is the basis for subsequent cognitive activity and learning. Interaction with the environment via sensory and motor functions generates trajectories whose relative similarity through the attractor maps is homomorphic with the structure of the perceived and enacted environmental interactions, with gradual modification of map structure and connectivity via dendritic and synaptic growth and atrophy.These trajectories represent the temporal structure of the environment.

- Churchland maintains that the vectors which constitute the attractor maps and trajectories through them constitute the brain's semantic system, which he refers to as 'state space semantics'. (Churchland 2012, 82).

Laakso & Cottrell (2000) proposed a modification of the Churchlands' position whereby the theoretical representational primitive should be partitioning of the high-dimensional neural state space into activation subspaces, replacing absolute position in the space: the number of units in the neural model determines the dimensionality of the state space and representations are vectors in the space, but because similar inputs cause similar vectors, activation - clusters form, and these clusters constitute the partition; the motivation was to counter critcism of state space semantics by Fodor & Lepore. This idea has subsequently been refined by Shea (2012, 2014, 2018), who argues that clusters in state space are the brain's 'vehicles of content'.

Conclusion

The foregoing historical outline shows that many serious thinkers over a very long time span have subscribed to the ideas with which this discussion began: that the meanings of linguistic expressions are their significations of mental concepts, and that concepts are representations of the mind-external environment generated by the cognitive agent's interaction with that environment. What the outline has added to this general position is:

- The distinction between natural and conventional signification, which aligns well with the original / derived intentionality one.
- The indentification of the neural mechanisms of environmentally-caused conceptual representation.

These ideas provide a plausible context in which models of intrinsic linguistic meaning can be constructed.

References

Adams, F., Aizawa, K. (2017) Causal theories of mental content, Stanford Encyclopedia of Philosophy, (Summer 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2017/entries/content-causal/>

Allan, K. (2009) The Western Classical Tradition in Linguistics, 2nd ed., Equinox Publishing

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain .Behavioral and Brain Sciences 33, 245–266

Atkin, A. (2010) Peirce's theory of signs, Stanford Encyclopedia of Philosophy, (Summer 2013 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2013/entries/peirce-semiotics/>

Barsalou, L. (2008) Grounded Cognition, *Annual Review of Psychology* 59, 617–645

Barsalou L. (2010) Grounded cognition: past, present, and future, *Topics in Cognitive Sciience 2*, 716– 724

Barsalou, L., (2016a) On Staying Grounded and Avoiding Quixotic Dead Ends, Psychonomic Bulletin & Review (2016) 23:1122–1142

Barsalou L (2016b) Situated conceptualization: Theory and applications. In: Foundations of embodied cognition: Volume 1 Perceptual and emotional embodiment*,* East Sussex: Psychology Press, 11–37

Barsalou, L. W. (2017) What does semantic tiling of the cortex tell us about semantics? Neuropsychologia, 105, pp. 18-38

Barsalou, L, Dutriaux, L., Scheepers, C. (2018) Moving beyond the distinction between concrete and abstract concepts, Philosophical Transactions of the Royal Society B: Biological Sciences 373, **PAGES**

Binder, J., Westbury, C., McKiernan, K., Possing, E., Medler, D. (2005) Distinct brain systems for processing concrete and abstract concepts, Journal of Cognitive Neuroscience 17, 905–917

Binder,J., Desai,R., Graves,W., Conant, L. (2009) Where Is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies, Cerebral Cortex, 19, 2767–2796

Binder J. (2016) In defense of abstract conceptual representations, Psychonomic Bulletin & Review 23, 1096–1108

Binder, J., Conant, L., Humphries, C., Fernandino, L., Simons, S., Aguilar, M., Desai, R.(2016) Toward a brain-based componential semantic representation, Cognitive Neuropsychology, 1–45.

Boeckx, C., Grohmann, K. (Eds.) (2013) The Cambridge handbook of biolinguistics, Cambridge, UK: Cambridge University Press

Bookheimer, S. (2002) Functional MRI of language: New approaches to understanding the cortical organization of semantic processing, Annual Review of Neuroscience 25, 151-188

Boone, W., Piccinini, G. (2016) The cognitive neuroscience revolution, Synthese 193, 1509-1534

Bradie, M., Harms, W. (2016) Evolutionary Epistemology, Stanford Encyclopedia of Philosophy, (Spring 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/epistemology-evolutionary/>

Buckner, C., Garson, J. (2019) Connectionism, Stanford Encyclopedia of Philosophy, (Fall 2019 Edition), ed. E. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/connectionism/>

Buroker, J. (2014) Port Royal logic, Stanford Encyclopedia of Philosophy, (Spring 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/port-royal-logic/>

Buss, D., 2007, Evolutionary Psychology: The New Science of the Mind, Boston: Allyn and Bacon

Carey, S. (2009). *The Origin of Concepts*, Oxford: Oxford University Press.

Churchland, P. (2012) Plato's camera. How the physical brain captures a landscape of abstract universals, Cambridge: MIT Press

Clark, A.. 2008. Supersizing the Mind: Embodiment, Action, and Cognitive Extension, New York: Oxford University Press.

Cohen, S. (2016) Aristotle's metaphysics, Stanford Encyclopedia of Philosophy, (Winter 2016 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/win2016/entries/aristotle-metaphysics/>

Cole, D. (2020) The Chinese Room Argument, *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), ed. E. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/chinese-room/>

Comesana, J., Klein, P. (2019) Skepticism, Stanford Encyclopedia of Philosophy, (Winter 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/win2019/entries/skepticism/>

Conway C., Pisoni D. (2008) Neurocognitive basis of implicit learning of sequential structure and its relation to language processing, Annals of the New York Academy of Sciences 1145, 113-31

Croft, W., Cruse, D. (2004) Cognitive linguistics, Cambridge: Cambridge University Press.

Crane, T., French, C. (2015) The Problem of Perception, Stanford Encyclopedia of Philosophy, (Spring 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/perception-problem/>

Downes, S. (2018) Evolutionary psychology, Stanford Encyclopedia of Philosophy, (Fall 2018 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/fall2018/entries/evolutionary-psychology/>

Fodor, J. (1975). The Language of Thought, Cambridge, MA: Harvard University Press

Fodor, J., and Lepore, E., 1996, Churchland on state space semantics, In Robert N. McCauley (ed.), The Churchlands and Their Critics. Blackwell. pp. 145--158

Fodor, J., and Lepore, E., 1999, "All at Sea in Semantic Space: Churchland on Meaning Similarity," *Journal of Philosophy*, 96: 381–403.

Friederici, A. (2011) The brain basis of language processing: from structure to function, Physiological Review 91, 1357-1392

Gazzaniga, M., Ivry, R., Mangun, G. (2019) Cognitive neuroscience: the biology of the mind, New York: W. W. Norton.

Geeraerts, D., Cuyckens, H. (2010) (eds.) The Oxford Handbook of Cognitive Linguistics, Oxford: Oxford University Press

Gendler Szabo, Z. (1998) Language, early modern philosophy of, Routledge Encyclopedia of Philosophy. Philosophy of Language, London and New York: Routledge, 128-36

Griffiths, P. (2009) The Distinction Between Innate and Acquired Characteristics, Stanford Encyclopedia of Philosophy, (Spring 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/innate-acquired/>

Harris J., Petersen, R., Diamond M. (2001) The cortical distribution of sensory memories. Neuron 30, 315–318

Jackendoff, R. (2002) Foundations of language: brain, meaning, grammar, evolution, Oxford: Oxford Univerity Press

Jackendoff, R. (2015) A user's guide to thought and meaning, Oxford: Oxford University Press

Jacob, P. (2019) Intentionality, Stanford Encyclopedia of Philosophy, (Spring 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2019/entries/intentionality/>

Kemmerer, D. (2015) Cognitive Neuroscience of Language, Psychology Press

King, P. (2007) Abelard on mental language, American Catholic Philosophical Quarterly 81 (2):169-187

King, P. & Arlig, A. (2010) Peter Abelard, Stanford Encyclopedia of Philosophy, (Fall 2018 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/fall2018/entries/abelard/>

Klima, G. (2017) The medieval problem of universals, Stanford Encyclopedia of Philosophy, (Winter 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/win2017/entries/universals-medieval/>

Laakso, A., and Cottrell, G., (2000) "Content and Cluster Analysis: Assessing Representational Similarity in Neural Systems," *Philosophical Psychology*, 13: 47–76

Lagerlund, H. (2017) Mental representations in medieval philosophy, Stanford Encyclopedia of Philosohy, (Fall 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/fall2017/entries/representation-medieval/>

Lakoff, G., & Johnson, M. (1999) *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*, New York: Basic Books

Langacker, R. (2008) Cognitive Grammar: A Basic Introduction. New York: Oxford University Press

Lau, J., Deutsch, M. (2014) Externalism about mental content, Stanford Encyclopedia of Philosophy, (Fall 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/fall2019/entries/content-externalism/>

Lenneberg, E. (1967) Biological foundations of language, New York: John Wiley and Sons

Levin, J. (2018) Functionalism, Stanford Encyclopedia of Philosophy, (Fall 2018 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/fall2018/entries/functionalism/>

Lycan, W. (2018) Philosophy of language: a contemporary perspective, 3rd ed., London: Routledge

Lyons, J. (2016) Epistemological Problems of Perception, Stanford Encyclopedia of Philosophy, (Spring 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/perception-episprob/>

Macleod, C. (2016) John Stuart Mill, Stanford Encyclopedia of Philosophy, (Fall 2018 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/fall2018/entries/mill/>

Marenbon, J. (2016) Anicius Manlius Severinus Boethius, Stanford Encyclopedia of Philosophy, (Winter 2016 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/win2016/entries/boethius/>

Margolis, E., Laurence, S.(2015). *The Conceptual Mind: New Directions in the Study of Concepts*, Cambridge, MA: MIT Press.

Margolis, E., Laurence, S. (2019) Concepts, Stanford Encyclopedia of Philosophy, (Summer 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2019/entries/concepts/>

Markie, P. (2017) Rationalism vs. Empircism, Stanford Encyclopedia of Philosophy, (Fall 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/fall2017/entries/rationalism-empiricism/>

Matheson, H., Barsalou, L. (2018) Embodiment and grounding in cognitive neuroscience. In: Wixted, J. (ed.) *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience [4th edition].* Wiley

McCawley, J. (1976) *Grammar and meaning*. New York: Academic Press

McCawley, J. (1995) Generative semantics, in: Verschuren et al. (eds.), *Handbook of Pragmatics* (Amsterdam & Philadelphia: Benjamins), 311-19

McKinsey, M. (2018) Skepticism and Content Externalism, Stanford Encyclopedia of Philosophy, (Summer 2018 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2018/entries/skepticism-content-externalism/>

Meier-Oeser, S. (2011) Medieval semiotics, Stanford Encyclopedia of Philosophy, (Summer 2011 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2011/entries/semiotics-medieval/>

Metzger M. (2017) Dynamical models of cognition, in: Magnani L., Bertolotti T. (eds) Springer Handbook of Model-Based Science. Springer Handbooks. Springer, Cham

Modrak, D. (2001) Aristotle's theory of language and meaning, Cambridge University Press

Neander, K. (2012) Teleological Theories of Mental Content, Stanford Encyclopedia of Philosophy, (Spring 2018 Edition) ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2018/entries/content-teleological/>

Neander, K. (2017) A mark of the mental: In defense of informational teleosemantics, Cambridge: MIT Press

Newmeyer, F. (1995) Generative linguistics: an historical perspective, London: Routledge

Oppy, G., Dowe, D. (2016) The Turing test, *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2019/entries/turing-test/>

Orzack, S., Forber, P. (2010) Adaptationism, Stanford Encyclopedia of Philosophy, (Spring 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/adaptationism/>

Patton, L. (2018) Hermann von Helmholtz, Stanford Encyclopedia of Philosophy, (Winter 2018 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/win2018/entries/hermann-helmholtz/>

Petersson, K., Hagoort, P. (2012) The neurobiology of syntax: beyond string sets, Philosophical Transactions of the Royal Society B 367, 1971-1983

Piccinini, G. (2007) Computationalism, the Church-Turing thesis, and the Church-Turing fallacy, Synthese 154, 97-120

Piccinini, G., (2009) Computationalism in the Philosophy of Mind, *Philosophy Compass* 4, 515–532

Piccinini, G., Scarantino, A. (2011) Information processing, computation, and cognition, Journal of Biological Physics 37, 1-38

Piccinini, G., Bahar, S. (2013) Neural computation and the computational theory of cognition, *Cognitive Science,* 34*,* 453–488

Piccinini, G., Shagrir, O. (2014) Foundations of computational neuroscience, Current Opinion in Neurobiology 25, 25-30

Piccinini, G. (2015) Physical computation: A mechanistic account, Oxford: Oxford University Press.

Piccinini, G. (2016) The computational theory of cognition, in Fundamental issues in artificial intelligence, ed. V. Müller, Switzerland: Springer

Piccinini, G. (2018) Computation and Representation in Cognitive Neuroscience, Minds and Machines 28, 1-6

Pinker, S. (2007). *The Stuff of Thought: Language as a Window into Human Nature.* London: Penguin.

Pitt, David (2012), "Mental Representation", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2020/entries/mental-representation/>

Pojman, P. (2019) Ernst Mach, Stanford Encyclopedia of Philosophy, (Spring 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2019/entries/ernst-mach/>

Port, R., van Gelder, T. (1995) Mind as motion: dynamics, behavior, and cognition, Cambridge: MIT Press

Prinz, J. (2002), Furnishing the Mind. Cambridge, Cambridge: MIT Press

Rescorla, M. (2020) The computational theory of mind, *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2020/entries/computational-mind/>

Rodriguez-Pereyra, G. (2014) Nominalism in metaphysics, Stanford Encyclopedia of Philosophy, (Summer 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2019/entries/nominalism-metaphysics/>

Rupert, R. (2008) "Causal Theories of Mental Content," Philosophy Compass, 3, 353–80

Ryder, D., 2004, "SINBAD Neurosemantics: A Theory of Mental Representation," *Mind & Language*, 19: 211–240

Rysiew, P. (2016) Naturalism in epistemology, Stanford Encyclopedia of Philosopy, (Spring 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/epistemology-naturalized/>

Samet, J. (2019) The Historical Controversies Surrounding Innateness, Stanford Encyclopedia of Philosopy, (Summer 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2019/entries/innateness-history/>

Samet, J., Zaitchik, D. (2017) Innateness and Contemporary Theories of Cognition, Stanford Encyclopedia of Philosophy(Fall 2017 Edition), ed. E. Zalta, URL =

<https://plato.stanford.edu/archives/fall2017/entries/innateness-cognition/>

Schmidhuber, J. (2015) Deep Learning in Neural Networks: An Overview. Neural Networks 61, 85-117

Searle, J. (1980) Minds, Brains and Programs, Behavioral and Brain Sciences, 3, 417–57

Searle, J. (1984) Minds, Brains and Science, Cambridge, MA: Harvard University Press.

Searle, J. (1989) ‘Artificial Intelligence and the Chinese Room: An Exchange’, *New York Review of Books*, 36: 2 (February 16, 1989).

Searle, J. (1990) ‘Is the Brain's Mind a Computer Program?’, Scientific American, 262(1): 26–31.

Searle, J. (1999) ‘The Chinese Room’, in R.A. Wilson and F. Keil (eds.), The MIT Encyclopedia of the Cognitive Sciences, Cambridge, MA: MIT Press.

Searle, J. (2002) ‘Twenty-one Years in the Chinese Room’, in Preston and Bishop (eds.) 2002, 51–69.

Searle, J. (2010) ‘ Why Dualism (and Materialism) Fail to Account for Consciousness ’ in Lee, Richard E. (ed) Questioning Nineteenth Century Assumptions about Knowledge, III: Dualism. NY: SUNY Press

Shagrir, O. (2012) Structural representations and the brain, The British Journal for the Philosophy of Science 63, 519-545.

Shea, Nicholas., 2012. “New thinking, innateness and inherited representation”. Phil. Trans. R. Soc. B, 367(1599): 2234–2244.

Shea, N. (2014). Exploitable isomorphism and structural representation. Proceedings of the Aristotelian Society*, 114,* 123–144

Shea, N. (2018) Representation in cognitive science, Oxford: Oxford University Press

Shields, C., The Oxford Handbook on Aristotle, Oxford: Oxford University Press, 2012

Shields, C., *Aristotle,* 2nd edition, London: Routledge, 2014.

Shields, C. (2015) Aristotle, Stanford Encyclopedia of Philosophy, (Winter 2016 Edition), ed. E. Zalta URL = <https://plato.stanford.edu/archives/win2016/entries/aristotle/>

Shields, C. (2016) Aristotle's psychology, Stanford Encyclopedia of Philosophy, (Winter 2016 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/win2016/entries/aristotle-psychology/>

Siegel, S. (2016) The Contents of Perception, Stanford Encyclopedia of Philosophy, (Winter 2016 Edition), ed. E. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/perception-contents/>

Spade, P. (2007) Thoughts, Words and Things: An Introduction to Late Mediaeval Logic and Semantic Theory, https://www.scribd.com/document/30475288/Spade-Thoughts-Words-and-Things

Spade, P. (2016) Medieval philosophy, Stanford Encyclopedia of Philosophy, (Summer 2018 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/sum2018/entries/medieval-philosophy/>

Spade, V. & Panaccio, C. (2019) William of Ockham, Stanford Encyclopedia of Philosophy, (Spring 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2019/entries/ockham/>

Stampe, D., 1977, “Toward a Causal Theory of Linguistic Representation,” in P. French, H. K. Wettstein, and T. E. Uehling (eds.), Midwest Studies in Philosophy, vol. 2, Minneapolis: University of Minnesota Press, pp. 42–63

Talmy, L. (2000) Toward a cognitive semantics, 2 vols., Cambridge: MIT Press

Thomson, E., Piccinini, G. (2018) Neural representations observed, Minds and Machines 28, 191-235

Tornau, C. (2019) Saint Augustine, Stanford Encyclopedia of Philosophy, (Winter 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/win2019/entries/augustine/>

Uzgalis, W. (2018) John Locke, Stanford Encyclopedia of Philosophy, (Spring 2019 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2019/entries/locke/>

Ward, L. (2001) Dynamical cognitive science, Cambridge: MIT Press

Wilson, R., Foglia, L. (2015) Embodied Cognition, Stanford Encyclopedia of Philosophy, (Spring 2017 Edition), ed. E. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/>

Wilson-Mendenhall C., Simmons W., Martin A., Barsalou L. (2013) Contextual processing of abstract concepts reveals neural representations of nonlinguistic semantic content. Journal of Cognitive Neuroscience 25, 920–35

Yeh, W., Barsalou, L. (2006) The situated nature of concepts. American Journal of Psychology, 119(3), pp.

349-384