University of São Paulo

Polytechnic School of Engineering

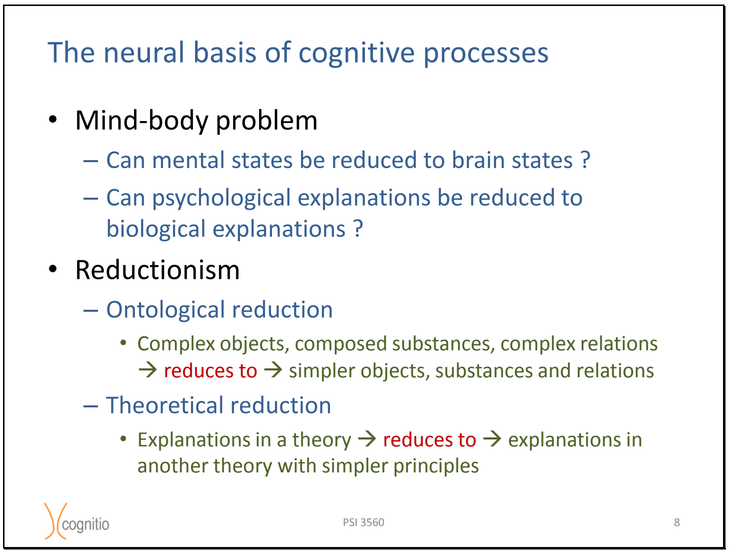
Department of Electronic Systems

Handouts of slides of class F5 – Modelling Cognition – Part II

Date: April 18, 2020 Author: João E. Kogler Jr.

This is the sequence of the material on part 1. I expect that you read this material after you have done and sent to me the exercises of part 1, and prepare for the remote discussion session, on April 28, via teleconference. Enjoy it and keep yourself safe and healthy.

Slide 8

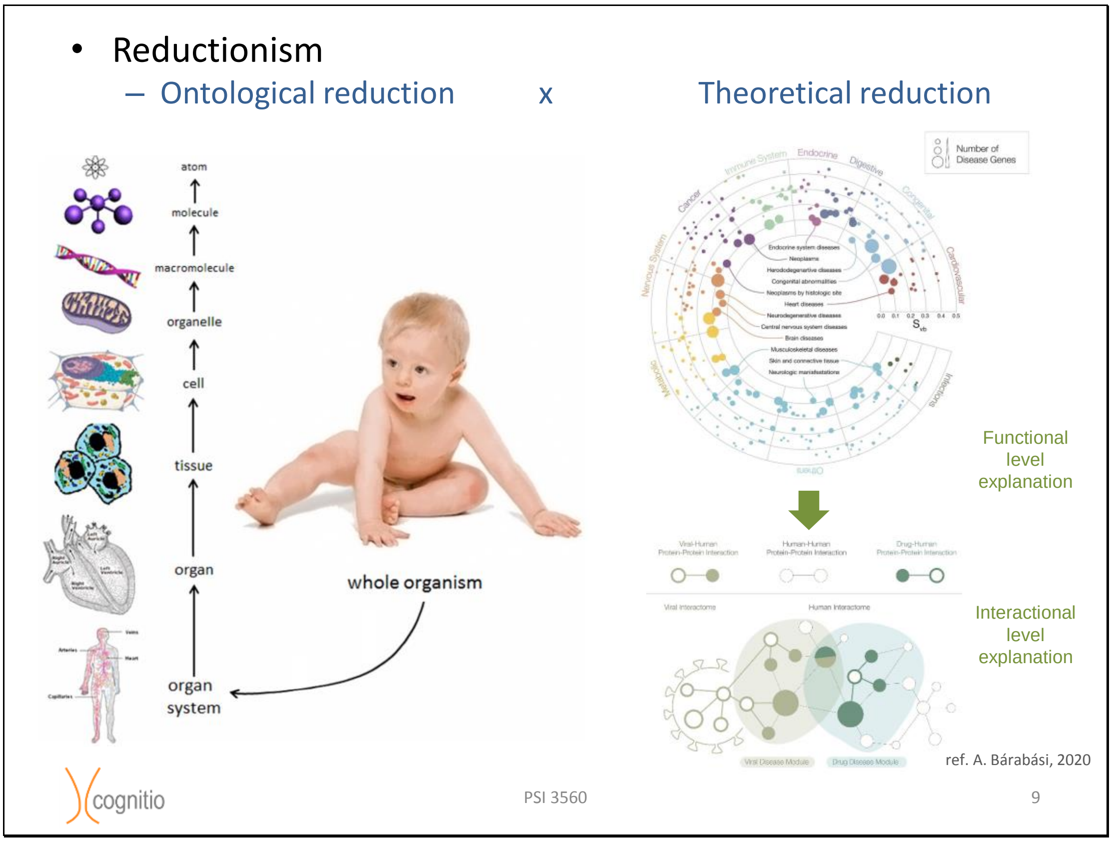


As stated on part 1 of F5, the central problem of cognitive science and philosophy of mind is to look for an answer to the question of reduction of the psychic to the physical. This translates in the two questions above depicted in slide 8, and one could ask if they constitute ontological or theoretical reductions. On either cases, they are reductions, which means that a more complex entity or explanation can be presented as a simpler one. Thus, the sense of the term "*to reduce*".

An ontology consists of a theory about the world, describing the objects, structures, properties and relations that compose it. Looking from the ontological point of view, the idea that the functioning of the brain, an organ, results from the collective behavior of its constituent parts (nervous tissue, neural circuits, neurons, glial cells, etc.) is an ontological reduction of organ physiology to cellular physiology. Also, the folk psychology view of the mind as constituted by thoughts, which are discourses composed of propositional attitudes, is an ontological reduction, of thoughts seen as mind discourses to the components of propositional attitudes.

In the mind-brain problem we have two distinct theories, one of psychological character and the other a biological one. Each theory has several levels of explanation, and separately each can be considered ontological reductions, as shown in the previous paragraph. However, a theoretical reduction of *the problem* relating mind and brain requires that the explanation in psychological terms shows logical correspondence with the explanation in terms of biological facts, i.e. that the theory of the mind reduces to a theory of the brain in all terms.

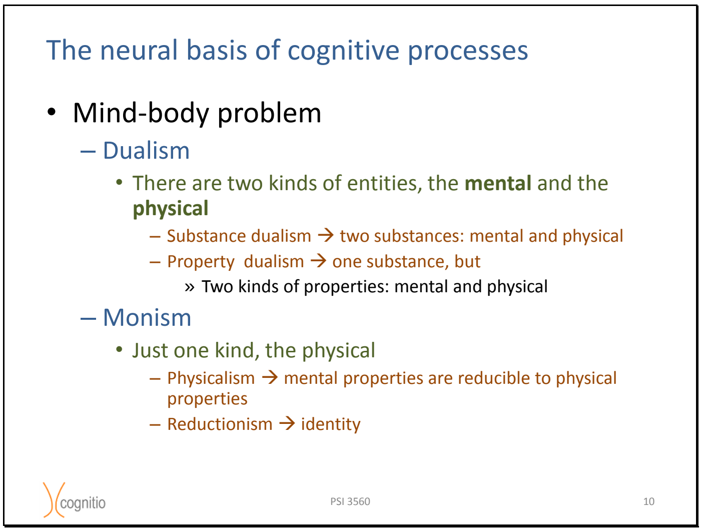
Slide 9



Before proceeding, let us show another example, of comparison of between ontological and theoretical reductions. The complete human body, as an organism, decomposes in organ systems. Each organ reduces to its constituent cells, which on their own are systems of cellular organelles, decomposable to the molecular level. Under this view, one can say that biology reduces to chemistry, which is reducible to physics. This reduction is called physicalism, an explanation of the world in terms of physical entities and phenomena, an ontological reduction. However, although this ontological reduction holds, it does not imply that biological theories are explainable in terms of physical theories: there is an explanatory gap between the two worlds, missing the explanation in physical terms of how the biological organization took place. Attempts to explain it did not go beyond allegations like “it emerges”, or “it self-organizes”.

The second example, of theoretical reduction, illustrates the current attempts done in research of virology, where the explanation of the mechanism of a viral disease put in terms of relations among several functional aspects like metabolism, immunity, genetic expression, systemic pathologies, etc., is paired with the explanation in terms of protein interactions in complex networks. These are two theoretic domains of explanations, namely biological and mathematical, and what is desired here is to reduce the former to the later.

Slide 10



As mentioned on last page, physicalism is the view that biological phenomena reduces to physical phenomena. We say that biological entities are constituted by physical entities, and that biological properties supervene on physical properties. Originally physicalism was designated as materialism, before the reduction of matter to energy on contemporary physics. This kind of physicalism is called object-based, because of its ontological character. Theory-based physicalism is the theoretical reduction to physical theories. In the case of the mind-brain problem, physicalism corresponds to the explanation of mind phenomena by means of physical theories. However, as there is who accept the physicalist reduction of mind, others are against it. The related schools of thought in this concern are the dualism and the monism. The dualism can be of substance and of property, this later also called non-reductive physicalism.

- Monism – for monists, the mind is brain activity, thoughts are illusory, mind is physical. For monism, there is only one type of substance – physical – and one type of property, physical too.
- Property dualism (a non-reductive physicalism) – for property dualists, mind comes from brain activity, but it has two kind of <u>properties</u>: physical activity (neural behavior) and psychical activity (thoughts), which is not reducible to the physical.
- Substance dualism – for substance dualists, there are two types of substances: the physical and the psychical (mental). The other case of dualism, the property dualism, proposes that there is just one type of substance – the physical – but it may present two types of properties: physical and psychical, as stated above. In contraposition.

Since early times in history, there were disagreements between philosophers. In Greece, Plato (427-347 BC) was a dualist and believed that mind existed in an ideal world of forms, <u>immaterial</u>

and non-extensible, while the body was in the material and extensible world. Aristotle (384-322 BC) Plato's disciple in his youth was, however, a monist. Aristotle proposed that mind is a property of the body, similar to the motion of its arms and legs, and that mind can be molded ("*moved*") by the body in the same way that clay can be molded to display different shapes. Nevertheless, neither Aristotle, nor Plato provided any theory about the (causal) interaction between body and mind. René Descartes (1596-1650), who started the so-called classical dualism, believed that there was a unidirectional causal link: the mind could control the body but not vice-versa. Descartes proposed that this one-way causal action was *via the pineal gland*.

There are several arguments against dualism. Besides the philosophical arguments, there are facts about mental perturbations that could be traced to cerebral disturbances, like those resulting from trauma, drug use and anatomic malformations. These facts suggest correlations between mental and brain features, which challenge the dualist view. On the other hand, dualists challenge the monist view by arguing that it is impossible to replicate in a physical system some attributes of mind, like consciousness for example.

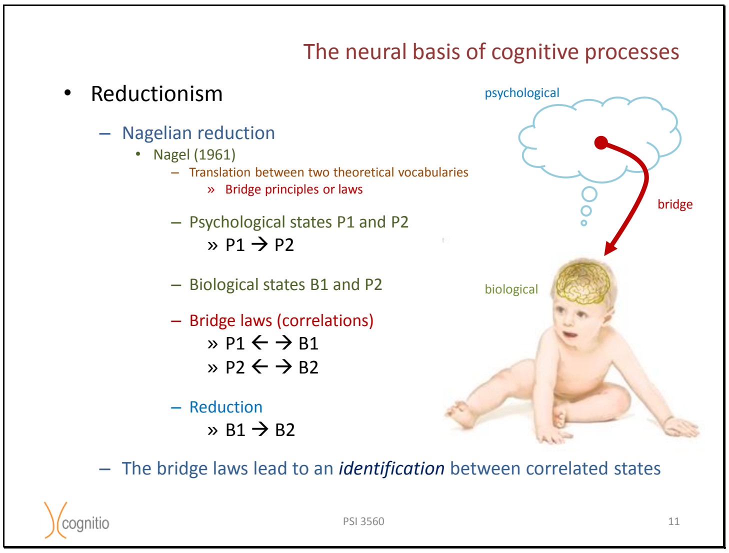
Nevertheless, just endorsing monism and proposing reductionism is not enough to solve the mind-brain question, because it does not explain how the reduction takes place. There is a corpus of knowledge provided by psychology as well as there is another provided by neuroscience, and the problem is how to put these corpora into correspondence?

The obvious answer consists in confronting them by comparison of laws and phenomena to stablish their correspondence. For this, one must know which laws of psychology pair to which laws of neuroscience and this is not simple, being problematic in at least in two aspects.

Firstly, in both theories the laws are not of paradigmatic theoretical type, like in is physics, where the genesis of a law starts with a theoretical hypothesis proposing some premise applicable to a wide class of phenomena, with its consequences and predictions. If experimentation can verify the predictions with favorable results, then the law is accepted. Differently from this, hypotheses in psychology or neuroscience refer to restricted experimental contexts. The coverage of these laws are of specific range in general, and do not result in universal laws, as in physics. In consequence, the pairing of psychological and biological laws must consider similarly restricted contexts and situations, for which is difficult to verify if the *ceteris paribus* conditions hold consistently to both. This is the methodological issue.

Secondly, the terms and concepts present in the psychological discourse are not precisely the same or may not be equivalent to those used by biology, or have equal interpretations. The lexical and semantic differences prompts a language gap. This is the ontological issue.

Slide 11



A scientific explanation, as proposed by Hempel & Oppenhein (1948), known as the covering-law view, satisfy three main items (see Weiskopf & Adams, 2015):

1. Scientific explanations are deductively valid arguments: the *explanandum* (the phenomenon to be explained) follows logically from the *explanans* (the explaining part):

   EXPLANANS → EXPLANANDUM

2. Among the premises in the *explanans*, there must be at least one verified natural law.
3. All premises in the *explanans* must be true.

Ernest Nagel (1901-1985) proposed that a scientific explanation can be deductively reduced to another by means of what he called bridge principle, or bridge law.

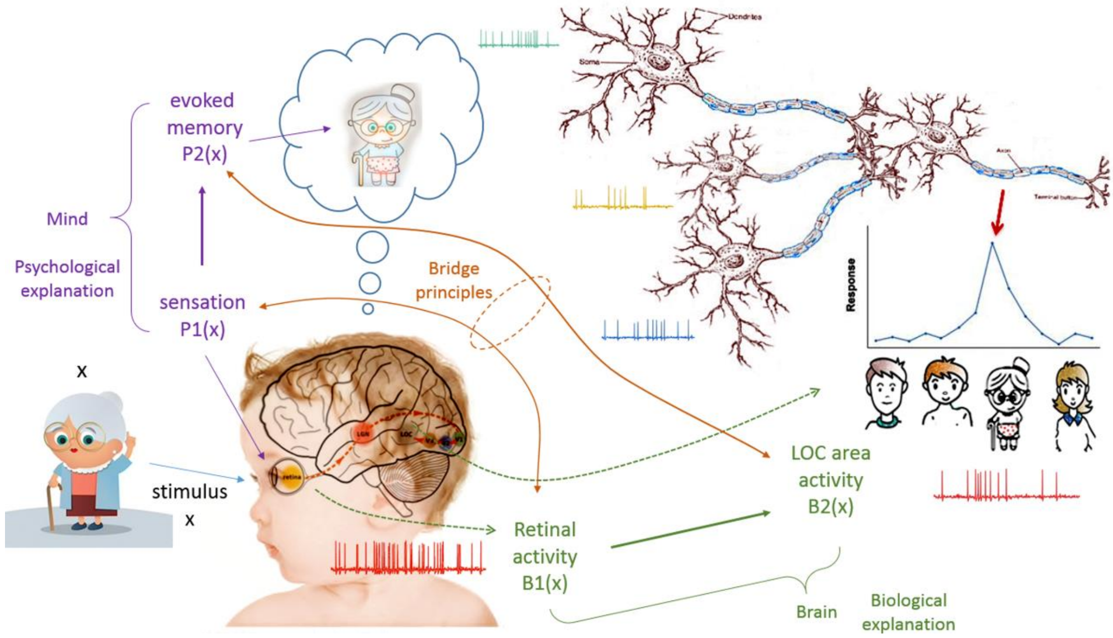
- Bridge laws are statements of the form T1(x) ←→T2(x) that put the terms on a theory T1 in correspondence with terms on another theory T2.
- They must translate an item of a theory into an item of the other, complementing the arguments on each side to compensate differences in vocabulary and expressiveness.

Consider a psychological law: **P1(x) → P2(x)** , which means that the phenomenon x has some feature explainable with a psychological explanation P2 (the psychological *explanandum*) that follows from a known psychological explanation P1 (the psychological *explanans*).

Now, suppose that some biological law refers to explanations about features of the same phenomenon x, however in biological terms: **B1(x) → B2(x)** , which means that biological explanation B2 (the biological *explanandum*) follows from a known biological explanation B1 (the biological *explanans*).

The reduction of P1(x) → P2(x) to B1(x) → B2(x) would then be obtained, according to Nagel, providing the bridge laws **P1(x) ←→ B1(x)** and **P2(x) ←→B2(x)**.

For instance, consider the following illustration:

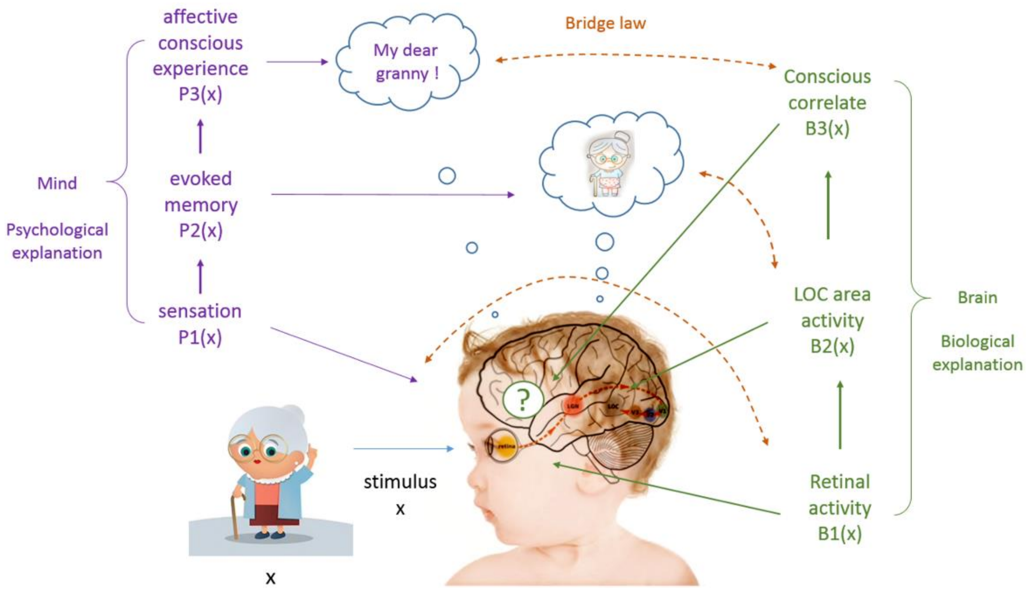


This situation describes the so-called "grandmother cell" principle: there are neuronal units in the cortical-ventral pathway area LOC (Lateral Occipital Complex) that maximally responds to some very specific visual stimuli, like the vision of your grandmother for instance. Therefore, the corresponding biological theory is that the retinal activity B1(x) in response to the visual stimulus x associated with the appearance of the grandmother, leads to a maximal activity B2(x) in the corresponding unit of LOC associated with the recognition of the grandmother (its response to faces of people other than the grandmother is significantly lower). A psychological theory of the vision of the grandmother would explain that the stimulus x provided by the appearance of the grandmother produces a sensation P1(x) by the visual system, which recalls a remembrance from episodic memory P2(x). The bridge principle applied to the corresponding parts of the theories B1(x) → B2(x) and P1(x) → P2(x) must deal with the inter-theory translations B1(x) ←→ P1(x) and B2(x) ←→ P2(x). If these translations are found, then the reduction is well succeeded. To obtain these translations it will be necessary to perform suitable experiments that show the correspondences between each mind state and their respective neural correlates, under compatible *ceteris paribus* conditions.

There is a curious consequence of the logical-deductive rationale behind the bridge principle. Because of this reduction, if some P2(x) → P3(x), then one could possibly find an explanation B3(x) in biological terms, by inferring a bridge P3(x) → B3(x). If one could show that B2(x) → B3(x), then the reduction would be consistent and P3(x) ← B3(x) follows.

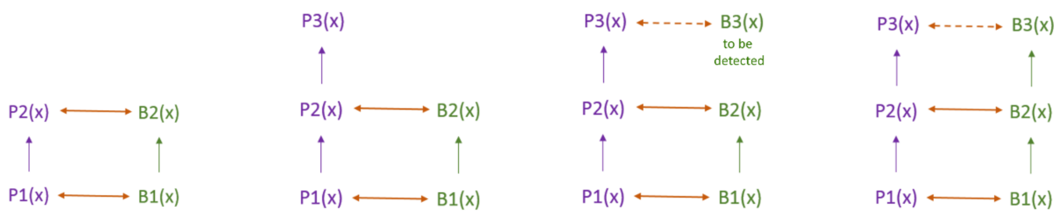
In other words, if a translation of some psychological explanations to biological explanations is stablished, further findings in psychological domain, derived from the previous ones, will meet biological explanations that can be verified as **logical** consequence of the previous ones.

In the previous example, suppose that P3(x) is an affective conscious experience related to grandmother's apparition, resulting from the evoked memory P2(x). Therefore, one would expect find a neural correlate B3(x) of P3(x) that should be traceable following a path from B2(x).

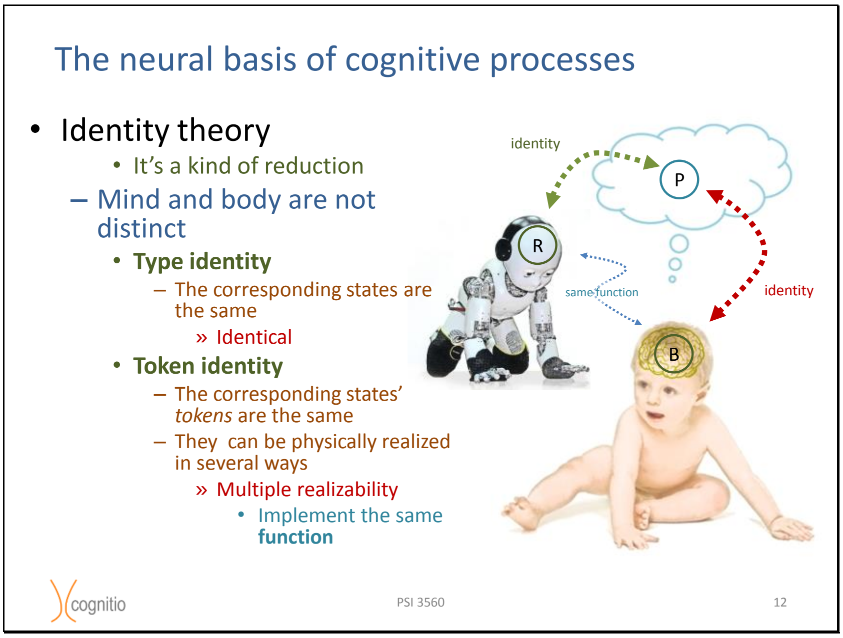


Remark:

- A point still to clarify is the sense of Nagelian reduction. All these constructs are logical-deductive ones. However, to show that B2(x) → B3(x) will require lots of experimental work, with statistical analysis providing the empirical confirmation of correlations. However, statistics on its own do not provide the required causal explanations: it just assesses correlations. Alternative hypotheses with suitable interventions and considerations of contrafactuals must form the basis of further experimentation, to verify the causal connection B2(x) → B3(x), which involve logical-deductive inference plus statistical induction. Therefore, although the process of reduction seems to be strictly logical-deductive, the whole work is not.



Slide 12



Still in the monist perspective, there is a more radical position than reductionism. Eliminative physicalism, also called eliminativism, rejects all explanations based on mental terms and claims that folk psychology should be dismissed in favor of biological explanations. For eliminativists, like Paul and Patricia Churchland (1981, 1986), mental states constitute a wrong concept because of its subjective character, and thus cannot be translated into objective brain states. For eliminative physicalism, only brain states should exist. Thus, there is no reduction, for the Churchlands[1].

For reductive physicalism, nevertheless, there are mental states <u>and</u> brain states, as concepts of distinct, but paired theories. However, this is the case of a theoretical reduction, and there still remains the question of what explains the existence of this pairing of theories. A ontological reduction is required, by reducing mental states to brains states, but this leads to a question of identity: are mental states and brain states *identical*, i.e. are them the same thing? For this question, the identity theorists offer <u>two points of view</u>:

- Type identity – D. Armstrong (1968), J. Smart (1959) and U. Place (1956) proposed that the mental states are the brain states. They are <u>the same</u> entity, i.e. are identical, have the same identity. Under this view, to have some psychological state, one <u>has to</u> possess <u>the strictly specific biological structure</u> that gives rise to it, because the psychological state <u>is the state</u> possessed by the biological structure. Psychological and biological

[1] However, we are interested on scientific psychology, instead of folk psychology, so eliminativism is out and we will follow reductive physicalism.

states are considered here ontologically identical. This is more than a theoretic reduction of one to the other: is equality, $P(x) = B(x)$. Therefore, bridge principles are not applicable here because the states are coincident [2], i.e. correlations $P(x) \leftrightarrow B(x)$ are not feasible.

- Token identity – a token is an indexical sign, i.e. an indicator of the entity to which it refers. For instance, the name of a person indicates the person without being the person itself: it just denotes the person's identity. H. Putnam (1967) and J. Fodor (1968), argued that type identity is troublesome, and proposed that mind states reducible to brain states have identical tokens: they represent the same thing, although this thing is not any of the denoted states, but some property or characteristic that both have: they do the same thing, have the same functionality.

For type identity, if someone feels pain (a psychological state), the person must have the specific and complete structure that enables one the feeling of that particular kind of pain. If something is missing in the structure, the person will not have that exact feeling of pain, not that psychological state. This is the sense by which type identity theorists require that a psychological state is identical to a biological state: for the biological structure to have that state, it must be complete, must be that specific structure capable of having that state.

Because of this sense, Putnam and Fodor presented several arguments that challenge type identity:

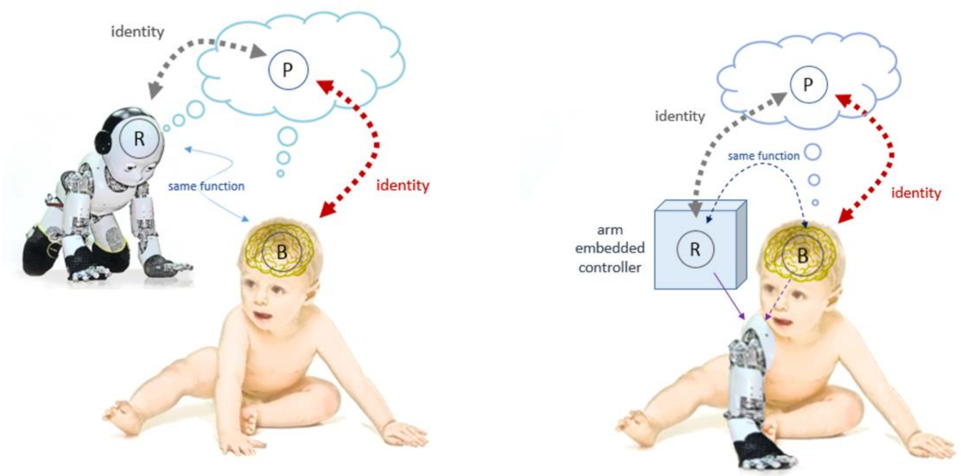
- Neuroplastic changes – neural associations can change without changing the meaning of the activity that the assembly of neurons displays. Take for example the case of phantom limbs: Ramachandran (1974) demonstrated that many people who lost a limb still present illusory perceptions as if they come from the "phantom" limb. This happened because the neural structures responsible for those percepts received signals from other parts of the body "invaded" that perceptual circuit. This result from neuroplasticity, as shown by Ramachandran (see, for instance, this following video).
- Convergent development – similarly to the case of the phantom limb, because of neuroplasticity it is likely that distinct neural structures can be correlated with the same psychological function in different persons. For instance, a person that learned French as her first language and some other that has it as his second language may have different neural structures performing the same function for speaking French.

[2] Notice that the position of type identity differs from that of eliminativism. For type identity theorists, mind states and brain states coincide identically, however one can still use mind terms. For eliminativists, mind states do not exist, therefor the mind terms must be eliminated from explanations.

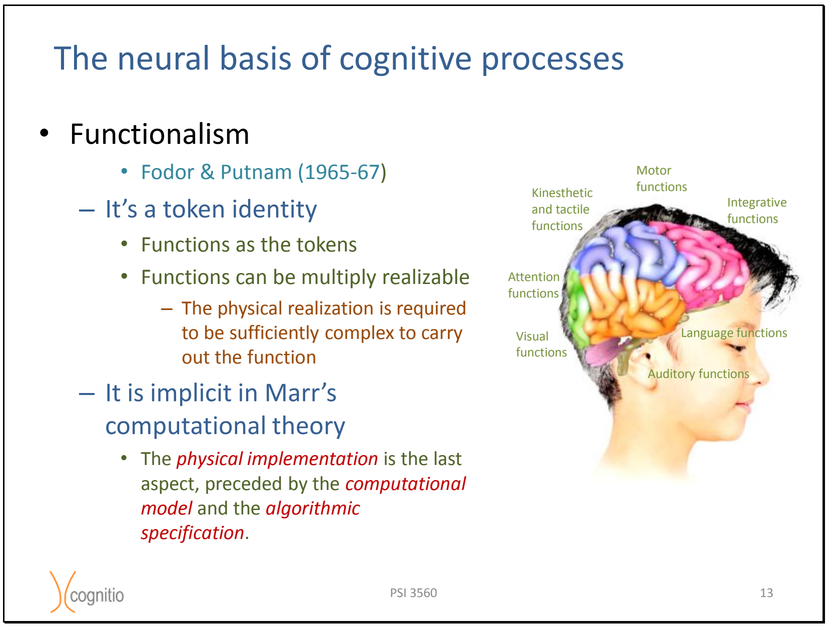
- Convergent evolution – distinct species may present different physical structures to produce the same psychological function, like perception, action, etc. as consequence of similar evolutionary pressure. For instance, in the case of visual perception, distinct species may have very different eyes and visual neural structures to produce similar perceptual functionality.
- Multiple realizability – the same psychological function can be physically realizable in very different ways, like biological versus artificial ones. If computers can perform tasks that humans can do, or distinct people or animals are able to produce the same psychological functions with distinct structures, these indicate that specific psychological states can correlate with distinct physical states of different physical structures.

These arguments point to the fact that psychological states and physical (biological) states are not the same, but denote the same functional entity: they correlate with the same function. Then, the tokens referred by token identity theory are the functions associated with the correlated states. The tokens translated into these functions are said to be identical. The neural state and/or structure which toke is identified with a psychological state's token is called the neural correlate of the psychological state.

In the slide picture, if a psychological state P(x) has a token TP(x) and its neural correlate B(x) has a token TB(x), then for token identity theory TP(x) = TB(x). Besides, invoking multiple realizability, if a robot computational state R(x) has a token TR(x) = TB(x), i.e. both the biological structure and the natural structure performs the same function, then TR(x) = TP(x) too. The function associated with all these states P(x), B(x) and R(x) is subsumed by the identical tokens TP(x) = TB(x) = TR(x). This fact is what serves as foundational theory for the operation of prostheses controlled by brain-machine interfaces (BMI).



Slide 13

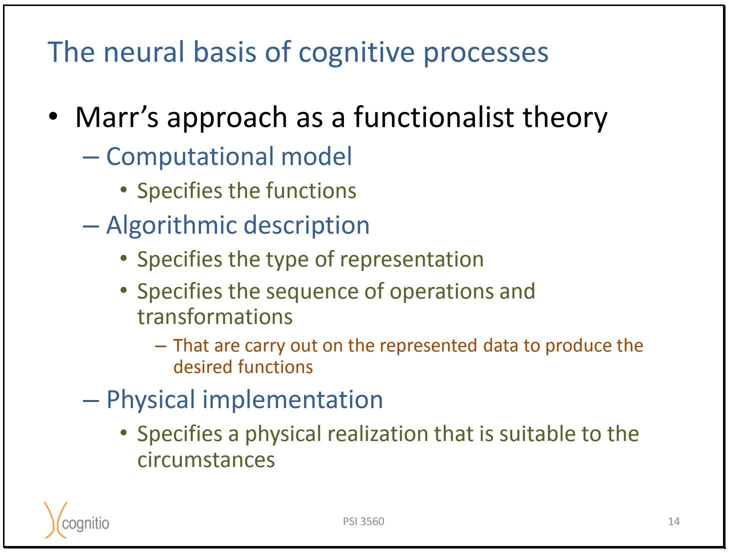


Based on the token identity idea, Putnam and Fodor proposed the functionalist approach to mind theory. Functionalism affirms that psychological states and processes are defined functionally, i.e. by the functions in which they are involved or perform. Consequently, psychological explanations can be stated in terms of these functions and they are reducible to biological, physical and computational explanations. The functions underlying these explanations are multiply realizable: it is possible to implement them on any kind of physical or formal system with organization, mechanisms and expressiveness coping with the functional complexity required.

An important consequence of the functionalist approach is that it entails the autonomy of psychology as science when related to other disciplines, in explaining psychic phenomena, because all psychological representations and their transformations will be tokens identical to those ones in the corresponding discourses of the other disciplines and levels of explanation.

It is in this sense that Marr suggested that a theoretical approach should precede the empirical exploration of new insights about the structure of brain internal activity, however exchanging the psychological explanations by computational theoretical explanations that would serve as a guide for finding neural correlates of mental activity. However, previously to it, the psychological explanations should then be translated into computational theoretical explanations, taking the tokens as bridge principles mapping the psychological terms into the computational ones. This factor is critical to make possible the idea of conceptual design as the first step considered in the development of a cognitive system.

Slide 14



The functionalist perspective is implicit Marr's approach to understanding perception, which we extended to the case of cognition. Since the functionalist view asserts multiple realizability, one can apply Marr's approach both to the investigation of natural living agents as for the design of artificial cognitive systems. It provides a way for functionally decomposing a system in differing levels of organization and complexity, performing thus a methodological reduction:

- Top-down decomposition in coarse-to-fine scales of complexity of functionalities present in a given system to reach a level of clear understanding.
- Bottom-up composition of fine-to-coarse scales of complexity (or low-level to high-level of abstraction) of functions, to reduce the conceptual gap between de desired functional capacity of the system and the functional capacities attainable by the components.

On the other hand, Marr's theory is an account of the ways to analyze the cognitive behavior of a living agent or the functioning of a sophisticated abstract process handled by an artificial agent or computational tool carrying a complex information-processing task. Marr's levels of analysis are (Marr, 1982):

- Computational theory – what is the goal of the computation, why is it appropriate, and what is the logic strategy by which it can be carried out?
- Representation an algorithm – How can this computational theory be implemented? In particular, what is the representation for the input and the output, and what is the algorithm for the transformations?
- Physical implementation – how can the representation and the algorithm be realized physically?

Slide 15

This is all for today.