



## Emergent Undecidability is Decidable\*

RENDRA SUROSO

Dept. Cognitive Science, Bandung Fe Institute  
Jl Cemara 63 Bandung, Indonesia  
brs@cogsci.bandungfe.net

---

### Abstract

Mind-machine analogy under the name Computational Theory of Mind (CTM) is, so far, the most promising way to understand mind thoroughly as computer. Only that the conventional way of doing CTM is by ignoring the way it is implemented in a physical architecture. Provided that the implementation is done through a design such that the machine is capable of being autonomous, then anti-computationalism that blossoms from Gödel Theorem becomes less relevant to the main idea of understanding mind through computer because what seems to be undecidable in a computational mind is actually decidable\* for an observer external to it.

**Keywords:** Gödel Theorem, connectionism, knowledge representation, Computational Theory of Mind, autonomous machine

---

### 1. Introduction

Since René Descartes, people have started to speculate on the possibility or impossibility to understand mind as a physical machine. Since Kurt Gödel and Alan Turing, people started to look back at the foundation of physical machine: formal system. Since J.R. Lucas, continues to Roger Penrose a few decades later, people started to doubt the use of (mechanical) formal system for explaining mind. Abstract and impractical this enterprise might be, but the impossibility to understand mind in terms of machine has attracted many methodologically-well-established researchers, only that they usually get away with it by applying a quick and dirty trick: let the mechanistic mind by itself incomplete or inconsistency since what to pursue is a theory of commonsensical mind where undecidability is common.

The trick soon leads to troubles: how to differ this so-called commonsensical from that that is not while that that is not is a formal system, and; how to find the commonsensical mind implemented in a physical rather than purely mental machine. The former will push us to reconsider what mind is in terms of a machine – or perhaps, many simultaneously interacting machines – and how *our* own mind as a machine can recognize other machines that in many respects is the same with ours but different in some others. The second will leave us no other choice but to again reconcile the old time mind-body problem.

On the other hand, it is not a matter of trivialities when Turing, being the first one who proposed the imitation game (Turing, 1950), mentioned the possibility of learning machine, which according to Lucas (1961), differs from our initial comprehension or definition of what a machine is.





Shortly after Turing's proposal, another idea of metaphysics – and then epistemology – comes up to be a popular discussion among philosophers: multiple realizability and then special sciences. Not until early 1990s, along with the advancement of simulational computation, the concept of emergence emerges in another field of research called complexity science, a compound of procedures which admits many independent laws of nature which appears to be very promising for verifying multiple realizability thesis and special sciences, though not very directly absorbed into philosophy of mind and cognitive science in general.

From this overview, I would very like to again show the possibility of understanding – or perhaps, modeling – mind as machines, but in terms of autonomous or learning machine such that the limitation of formal system can be logically ignored and we then do not have to bother with Gödel Theorem but for non-trivial reasons.

Unfortunately, everything has its cost. To do this, we will have to take a closer look at and even to revise the mainstream of viewing mind as a syntactic-manipulating computational machine in section 2. In section 3, implemented syntactic-manipulating computational machine is again given a review that comes up with a possibility of 'mild' reductionism in virtue of emergent properties surrounding the issue of cognitive mind. Section 4 considers a new form of machine, autonomous or learning one so it becomes different from machine Lucas ever mentioned but sufficient to describe emergent properties. Lastly, in section 5, all preceding sections will give us new directions to view fundamental problems in computational mind, i.e.: the quick and dirty trick, and of course, undecidability.

## 2. Mind as a Machine

For now, I will not mention Turing's proposal on imitation game for understanding intelligence. Instead, I will review the definition of machine Turing uses in imitation game before we can trace back its potential pitfalls on behalf of Gödel Theorem or Turing's halting problem.

In Lucas' (1961:126) terms, mechanical machine required to model the mind follows 'mechanical principles' such that:

“...we can understand the operation of the whole in terms of the operations of its parts, and the operation of each part either shall be determined by its initial state and the construction of the machine, or shall be a random choice between a determinate number of determinate operations.”

Having reached this point, there is no use to turn back and escape the *analogy* between mind and computer.<sup>1</sup> Otherwise, the reader might just stop reading this paper. Computational mind brings us a 'classical' theory of computationalism, or in more philosophical – rather than computational – details, often called Computational Theory of Mind (CTM; Fodor, 2000) with these following principles:

1. Propositional attitudes (belief, desire, thought, and the like) have their causal roles in virtue of, *inter alia*, their logical form.
2. The logical form of a propositional attitude supervenes on the syntactic form of the corresponding mental representation.
3. Mental processes (including, paradigmatically, thinking), are computations, that is, they are operations defined in the syntax of mental representations, and they are reliably truth preserving in indefinitely many cases.

<sup>1</sup>The moral is still the same: that we are to start somewhere, though the analogy has its own cost such that symbol-grounding problem (Harnad, 1994), Chinese Room argument (Searle, 1980), problems of minds of other organisms and non-conspecifics (Nagel, 1974), and many more similar puzzles I would not like to pursue depict the risk very well.





These principles do not pass without difficulties. Propositional attitudes have puzzling contents or meanings (Chalmers, 2002). Syntactic propositional attitudes have constituents called concepts but some of them are found to be non-conceptual (Bermúdez, 1995). Even Fodor (2000) himself insists that the syntax of mind has more than just ‘internal’ symbols, though probably, this insistence is only to keep his syntactically-minded principles of CTM intact under the name of holism, whilst mental representation also requires a semantics. And so on.

We only should note that, taking mind computationally, albeit mere analogously, is the same with allowing mind to be truth-preserving syntactically-processing machine in ‘indefinitely many cases’. This is necessary condition for our Gödelian Achilles’ heel to be encoded (Lucas, 1998), though the term ‘indefinitely’ or ‘infinitely’ does not seem to completely be realized by a computational mind.

CTM, on the other hand, in more furnished detail is used to attack the idea of massive modularity thesis from evolutionary theorists and following this line of argument is outside the scope of the paper except some emphases. I will only particularly elaborate how the machine is implemented in a bio-mechanical body, or at least, what (physical) components it necessitates. These components are necessary and sufficient condition to explain how our specific kind of mind of *H. sapiens* works (and not every possible mind functionalists might suggest) in terms of mechanical syntactic processing that follows rules of formal system.

If mind is roughly a computer program, then stating mind as syntactic manipulation implemented in a biological body – whereas syntactic manipulation itself is deemed to be inconsistent or incomplete – will lead us to the problem of emergence. But first, before moving further to emergence, I will take a brief look at the problem of implementation.

**Definition 1** Mind is syntactic in formal language  $L$  which contains well-formed formulas corresponding to propositional attitudes towards concepts or composition of concepts  $x$ .

### 3. Problem of Implementation

*A symbol is a pattern, any pattern that denotes or points to some other pattern.*  
Herbert Simon (1995)

Computational mind has, since the genesis of cognitive science in the 1950s, at least two different methods of implementation, i.e., connectionism and knowledge representation, each won its own ups and downs.

#### 3.1. Connectionism

It is widely known that connectionist approach to explain cognitive mind is also through computation with neurophysiological laws as its explanans. Here, the term computational is not in sense that it explicitly follows syntactic manipulation, but still algorithmic one must admit since pattern recognition as its goal to be reached by specific network is done by elementary units or elementary networks, e.g.: perceptron (Bates & Elman, 1993). If syntactic manipulation is the pattern recognition itself, then what is wrong with connectionism?

Since 1980s, connectionism has won a great reputation for providing a more plausible or realistic (biological) machinery for a cognitive mind. The strength of the models connectionists construct is that the models are directly attached to every possible observed behavior and perceptual system, as much as to ‘central processing unit’ of mind in the central nervous system, assuming that we are made of functionally-similar and interconnected neurons through and through. The fundamental premise





is that individual neurons *do not transmit large amounts of symbolic information*. Instead they compute by being *appropriately connected* to large numbers of similar units (Feldman & Ballard, 1982).

The weakness is that the actual capability of the model is far away from what for so long has been considered to be cognitive. In other words, it is hard to imagine how a formal system is implemented in a connectionist model. It is unlikely that we can so easily construct a connectionist architecture such that it can function as, say, syntactically-adequate theorem prover. In reply, some extreme of connectionists says that connectionist models have their own 'evidential' logic in contrast to symbolic logic of conventional computing, i.e., an associationistically-driven logic in a world that is perfectly controlled by induction such that our ability to recognize formal system is only a result of, more or less, associationistic neural plasticity. Through this view, our Gödelian problem becomes moot, whereas it is not.

If taken to semantic consideration, notorious arguments from Fodor and Pylyshyn (1988) on a bunch of abstract constraints called productivity and systematicity denounce connectionist movement abruptly. The arguments are not polemic-free, of course, but it successfully reminds connectionist researchers that mind does not, for instance, separately operate in natural language, but a mental language containing propositional attitudes often called Language of Thought.

More or less, this Language of Thought itself is the language CTM proposes, and has been in very much accordance with mainstream AI ever since (e.g.: Saphiro, 2003). In general, with more elaborated axiomatization, this mainstream AI is popularly known as knowledge representation.

**Definition 2** Mind is pattern recognition of a finite set of vectors  $\{\xi^\mu\}_\mu$  such that syntactic manipulation of input-output pair  $(\xi_p, \xi_d)$  that corresponds to the pattern recognition, if any, is of the form  $y(\xi_p, \xi_d)$ .

### 3.2. Knowledge Representation

Knowledge representation approach is probably what CTM and Lucas' machine denote. Therefore, knowledge representation does not concern much on implementation. Everything that goes is completely mentalistic, irreducible to neurophysiological or any other underlying lower phenomena.

Furthermore, provided that irreducibility is true, purely moving in logical forms as the contents of mind, perhaps automatedly as what machines do, has not come to any safer place. Human does not always draw correct inferences, or more precisely, human does not derive proof of a theorem as an automated-reasoning machine does. Almost every artificial intelligence researcher feels comfortable to say that commonsensical thinking only involves very limited logical capabilities; the quick and dirty trick itself. Let us call it hereupon: *derivability problem*.

On the other hand, the fact that the borderline between what is commonsensical and what is not turns out to be bleak. What to be necessarily acquired into knowledge base becomes a serious problem, hence, the *frame problem* (McCarthy, Hayes, 1969).

Lastly, KR-based computational mind is where the Suyodhana's thigh, the Achilles' heel, the red herring of logic, the Gödel Theorem, resides.

**Definition 3** Mind is syntactic manipulation of input-output pair  $(x_p, x_d)$  through a derivation function  $f(x_p, x_d)$ .

### 3.3. Reconciliation

To reconcile both, there is no other choice but to assign some intermediate relations between both approaches. Usually, these intermediate relations are called emergence and downward causation, two things so commonly found in complexity science.





Suppose the machine now consists of two different levels: micro-level, connectionist model, and; macro-level, KR-based model. The fact that micro-level can emerge macro-level is, for an external observer, unpredictable. The machine becomes autonomous. This is where my next explanation should be next directed to.

But first, summary of this section:

- we should give account to implementation problem of computational mind in its underlying neurophysiological system so we can explain the property of computational mind based on its underlying neurophysiological level;
- connectionism can model neurophysiological phenomena adequately, but how it emerges syntactic computational mind remains unknown;
- knowledge representation as the model of computational mind has, inter alia, these properties: it must have all criteria formal system must have (soundness, completeness, consistency, etc.) in favor of syntactic properties of computational mind formal logic and CTM suggest, and; it must have abductive power such that there is something more than just syntactic properties in computational mind, that is, whatever relates one logical form to another<sup>2</sup> outside logical inference formal system is capable to perform.

**Proposition 1** For a mental representation, there is necessarily a relation of  $\gamma(\xi, \xi)$  and  $f(x_p, x)$ .

#### 4. Autonomous Machine

For a machine to be autonomous, first, we have to recognize its autonomy. Hence, we put ourselves to be the observers to the machine such that we can do anything to the machine except one thing: having it redesigned. If there is a response, then we, as the observers, can expect that the response is or is not achieved by the machine only based on particular stimuli and design containing axioms and rules we initially put in. This is necessary for the machine to be called autonomous.

Shortly speaking, autonomous machine is free to modify itself – acquire new propositions as those of new concepts, block some illegal inferences just like mind can rule out nonsense, find shorter path of proving a theorem in virtue of default reasoning; all summarized in a bundle of general principles.

In Turing's (1950) own words:

“One might try to make it as simple as possible consistently with the general principles. Alternatively one might have a complete system of logical inference “built in.””

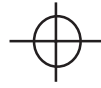
Pretty much in accordance with dominating psychology of his time, Turing called the former machine learning machine instead of autonomous – he used child-computer analogy a lot for that. And his second machine is our ordinary KR-based machine.

A regress argument contra reductionist: Even if this kind of autonomous machine is non-deterministic or not explicitly-algorithmic (Grush & Churchland, 1995), how it operates non-deterministically is previously determined, e.g.: stochastically, heuristically; with the help of a formal system the observer uses. We need to do better than this.

And apparently, the difference between autonomous and merely automated machine is only there in initial design, since both are initially designed with the help of formal system. For our next purpose, to differ them then we have to:

<sup>2</sup> Fodor's notion of 'globality' or 'holism' related to abduction is not very clear, contained in the idea of productivity and systematicity, not supported by empirical claims contrary to his and Pylyshyn's claim (van Gelder, Niklasson, 1994), and in an other occasion (Fodor, 1998), is said to do with 'semantic' – not syntactic – properties of logical forms.





1. bear the position as an external observer such that we can acknowledge through formal system external to the machine that the machine is autonomous or not. Without being an external observer, we do not know what our own initial design is, and we do not even know whether we are autonomous since we only suffer *shifits* – details to follow – that happen in our propositional attitude level, and;
2. also consider what the machine is made of instead of merely how they function.

#### 4.1. Subjectivity

Now why bothered with subjectivism in a cold-blooded scientific program? This section is primarily related to us being the observer, and not the machine that we observe.

As scientists, we always perceive, sense, think and make abstractions or fit matters into our own picture of scientific explanation before we finally generate laws of them. For all this, we do not have to look back to ontological issues such as old time idealism or more modern anti-realism for that I will adopt typical scientific realist view without comment.

Sufficiently speaking, there are scientific laws of nature or at least, regularities observed in many different levels, whether there are or are not us as the observers. What is not guaranteed by mere scientific realism is how, from the first place, we can determine from which laws (or axioms, or knowledge base, perhaps of nature) our theorems about particular events or occurrences or instances should be derived through a describing language – we probably owe a quotidian explanation to holism for this.

A lesson from systems theory of the 1950s shows us that this problem will evidently and so easily becomes a subject of subjectivism, and what is taken as truism is simply the way to construct, derive, solve, and make a pictorial representation of differential equations. Radical constructivism is a good example, such that Roger Sperry (1991:227), being a systems theorist himself, comes to a lame conclusion concerning mentalism in dynamical system that:

“In respect to this new legitimacy [of macromental causality] of the subjective in science it is important to remember that this does little or nothing to remedy the long-recognized methodological difficulties of dealing with introspective entities in science.”

In turn, taking mind as an autonomous machine is not as easy as allowing whatever kinds of knowledge base to be applied by our observing mind. There are some explicit constraints in doing so, and two of them must be that:

1. our cognitive mind only contains one macro-level, propositional attitude level such that we do not talk about us having an epileptic seizure in terms of our EEG being flat, but that we, inter alia, believe or think that our EEG is flat such that we suffer an epileptic seizure (after we get up) or think about nothing at all (during the seizure). Due to this single level character of our computational mind, there is no way to conclude that we are autonomous or not since autonomy requires more than one level, and;
2. our propositional attitude level is open for certain form of reduction to any (but not arbitrary) micro-level – the widely accepted but not convincingly proven is of course neurophysiological level which is computationally implemented in connectionist models – without dismissing existing propositional attitude level, or otherwise, our pictures of reality and cognitive mind will end up with irritating pluralism, or in different fashion, functionalism.

*We could be made of Swiss cheese and it wouldn't matter.*

Hilary Putnam (1975)





4.2. Emergence

Following 4.1. on subjectivity, hence I believe that putting physicalist-functionalist philosophical debate aside will leave my subsequent elaboration intact.

It is only that between micro-level and macro-level, between mental and biological, between propositional attitudes and neurophysiological phenomena, there are some laws which operate as the intermediate processes I previously mentioned in 3.3. In terms of propositional logic, Ernest Nagel (1961) called this laws ‘bridge laws’ (see Block, 1997 for discussion). In terms of continuum approach, Stephen Pepper (1926) called these laws ‘emergent laws’ that governs upward causation though the result is merely epiphenomenal. Furthermore, an instance in micro-level do not give us right to deduce an event in macro-level; we need shift (or same-level causation one may say), probably in both levels. The difference between discrete propositional logic and continuous algebraic picture might be of significance since it will assign continuous picture to micro-level (the brain works in action potentials) and discrete picture to macro-level (we think in propositions).

To precisely compare notes with Pepper, Nagel, and Paul Meehl and Wilfrid Sellars (1956), emergence is as depicted by Figure 1.

Pepper’s picture is an anti-reductionist picture of the world, while more furnished picture of Meehl and Sellars is one that makes reductionism possible. As has been said, to precisely verify which one is true between reductive-physicalism and functionalism is part of another job I will only review very briefly in the rest of the paper. But at least still I owe them the notations.

Sufficiently speaking, I will only take a characteristic of the so-called emergent property known as unpredictability.

**Proposition 2** Emergence is a condition such as, according to an external observer with formal system  $L^*$ , there is a shift from  $\xi_i$  to  $\xi_o$  that corresponds to an emergent property  $x$  on the observed formal system  $L$ .

4.3. Unpredictability

Following Meehl and Sellars’ picture of emergent phenomena, we must make the idea of unpredictable characteristics as unequivocal as possible, that is, as shown in Figure 1, unpredictability shows that there is only a shift from  $\xi_i$  to  $\xi_o$  that becomes necessary and sufficient condition for  $x$  to happen for all possible  $\xi$  as governed by bridge law  $g(\xi, \xi_o; x)$ , while other shift do not emerge the same  $x$  through the same bridge law. If the same  $x$  is emerged by different shifts, say,  $\xi_i^*$  to  $\xi_o^*$ , then we will

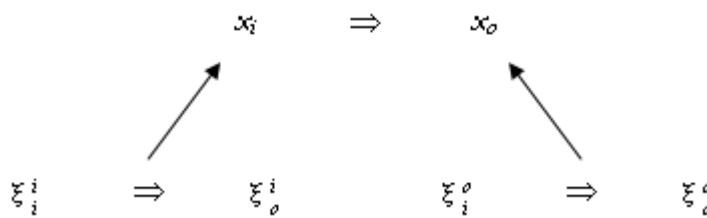


Figure 1 Emergence

For Pepper, a shift of lower level  $\xi_i$  to  $\xi_o$  is as observed, emerges upper level  $x$  such that this emergence is sheer epiphenomenon or mere quality. For Meehl and Sellars, such shift is a necessary and sufficient condition for an emergent property  $x$  such that there is a chance to explain all only in terms of  $\xi$ . For Nagel, the emergence is governed by bridge law  $g(\xi, \xi_o; x)$ .





meet multiple realizability condition.<sup>3</sup> On the other hand, the shift from  $\xi_i$  to  $\xi_o$  is predictable in sense that it obeys a particular law  $y(\xi_p, \xi)$ .

Now it is safe to propose an argument by analogy that, taking  $\xi$  as our neurophysiological level and  $x$  as propositional attitude level, we can deduce what varieties of  $x$  which will emerge with the help of particular bridge law  $g(\xi_p, \xi; x)$ , despite the fact that  $x$  is also governed by its own shift  $f(x_p, x)$ . The question is, are there such bridge laws?

Scientific optimism may translate this doubtful enterprise in terms of ‘very remote’ instead of ‘certainly not’ (Grush & Churchland, 1995). Probably it takes another millennium to make this sufficiently established – and such a positivistic bully it may sound. Therefore, for now, I will only sketch that propositional attitudes also follow other kinds of regularities, at least, merely empirically.

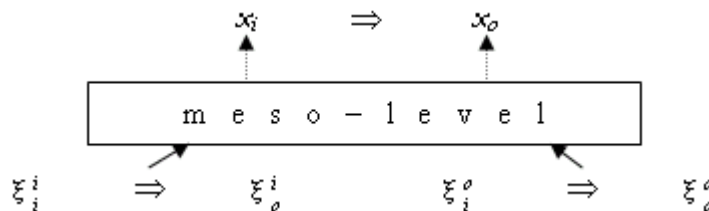
**Proposition 3** Unpredictability is a condition such that there is no bridge law  $g(\xi_p, \xi; x)$  that maps each shift from  $\xi_i$  to  $\xi_o$  for all possible  $\xi$ , to corresponding  $x$  for all possible  $x$ .

### 5. Mind as Machines

Empirical claims which are rooted in experiments in developmental psychology and modularity thesis lead us to say that mind is, at least, partially modular, or in previous works often called moderately massive modular (Carruthers, 2003; Suroso, 2004). The celebrated Wason Selection Task and Kahneman and Tversky’s risk aversion and risk seeking behaviors are just two wellknown results.

In concurrence with Peter Carruthers, I must say that mind, despite being autonomous, is engineeredly kludgy. What really strikes me at this point is that there must be some intermediate level, hence, meso-level.

With the help of this meso-level, then it is more plausible to think of upward and downward causation<sup>4</sup> at the same time. But what is this to do with meso-level of moderately massive modularity thesis?



**Figure 2 Meso-level of Cognition**

What we have from the existence of such meso-level as suggested by developmental psychology is that variables in meso-level are probably still emerged by micro-level, only that this meso-level does not necessarily emerge macro-level (indicated by dashed line), but instead, limit every possible  $x$  to enter mental representation or limit  $f$  such that not for every mental representation, truth-preserving condition is met. Hence,  $f$  becomes partially context-sensitive, though not necessarily massively modular.

<sup>3</sup> Multiple realizability is also possible for that there is another shift  $\delta_i$  to  $\delta_o$  such that  $\delta$  is completely different in properties from  $\xi$  but functionally isomorphic (Putnam, 1975) with  $\xi$  if put into the same functional properties  $\xi$  initially exhibit. This is another key idea of functionalism.

<sup>4</sup> In addition, concerning downward causation, old time wisdom says that the mental cannot move physical object, unless a behavior motivated by the mental successfully initiates a motori-movement such that the physical object is eventually moved. But then, it turns out that, provided that the central nervous system containing neurons usually modeled by connectionist is the most responsible for entire mental properties, downward causation to manipulate the chaotically-dynamic brain (Freeman, 2000) *directly* is less of importance because:

- the result of downward causation will give no significant change in micro-level due to its chaotic nature, as much as we cannot differ EEGs of our conscious thinking and relaxing states, or otherwise, for instance, suicide would be so easy, and;
- downward causation is, after all, mostly realized such that macro-level can generate motori-mechanism. If taken too far, we will finally end up with situated cognition for this, which I won't.







To fill in the blanks, now we have to characterize the meso-level properties such that Fig. 1 should be modified as shown in Figure 2.

If this meso-level is true, then it will be a new direction connectionist models should explain instead of typical problems such as constructing language acquisition machine or even Lisp-like processing machine with neurons as the building blocks; unless the models attempt to explain only sensori-motory mechanism classical modularity thesis suggests and limits – 2½D-sketch rendered to 3D-image is a good example for this.

Unfortunately, albeit very promising to reconcile connectionist and classicalist or physicalist and functionalist debates, we do not yet have further details such that the term ‘kludgy’ comes so easily. Fine-grained localization of brain areas through different functional levels performs very poorly if mapped onto coarse-grained mental phenomena (Bechtel & Mundale, 1999). Thus, it is worth speculating that (mental) concepts and inferences governed by folk-theoretical laws are processed and stored in this (mental) meso-level, before accessed by the global syntactic manipulation occurring in macro-level.

Thus it is sufficient to say that our subjective or observing nature is guaranteed by adding this level to entire architecture of computational mind.

**Proposition 4 Moderately Massive Modularity** Meso-level is emerged by micro-level, only that meso-level does not emerge macro-level. Instead, for an external observer, it applies  $f'(x_i, x_o)$  such that there exists some  $x$  that does not belong to either  $x_i$  or  $x_o$  for a function  $f$ , and  $f \neq f'$ .

### 5.1. On the Quick and Dirty Trick

Let us turn back to rationale of modern day artificial intelligence researches that face the familiar derivability problem and frame problem. Now, provided that if commonsensical reasoning is, at least mainly, governed by concepts and folk-theoretical laws, while concepts and folk-theoretical laws are context-sensitive, then Fodor’s problem of abduction vanishes, except for those mental representations not governed by folk-theoretical laws. In other words, in this meso-level, we are dealing with propositions, that is, concepts and folk-theoretical laws mixed compositionally but not contingently – because they are presumably innate.

Though not very clear and still (again) kludgy in practice, during our lifetime we can only recognize about 60.000 concepts (Laurence & Margolis, 2002). This might still be a huge number, but the task of AI researchers to cope with the frame problem and derivability problem hence becomes a bit simpler, at least in principle, provided that the corresponding folk-theoretical laws for each concept or sequence of concepts are established, and computational mind is not that context-insensitive anymore.

The quick and dirty trick from Marvin Minsky, John McCarthy, and particularly from Stuart Shapiro (1995) who says against Lucas as follows:

“There are many Gödel-type sentences, but they all involve Knights (who never lie), Knaves (who never tell the truth), and logicians (who, at least always reason logically), so it might be felt that these sentences have nothing to do with the normal human mind.”;

becomes less dirty although also less quick, that is, instead of banning the existence of sentences these Knights, Knaves and logicians may utter based on the knowledge base of the machine, it is allowed to ban or just let those Gödel-type sentences to actually exist, all by itself. Only that to do this, we also need to consider the micro-level and perhaps the meso-level.





**Lemma 1** In mind expressed in a language  $L$ , there possibly exists  $x$  and  $\neg x$  due to the fact that there are  $f(x_i, x)$  and  $f(x_i, \neg x)$ . But for an external observer in language  $L^*$ , there is no inconsistency due to the fact that in  $L^*$ , there is a derivation function that operates through meso-level  $f^*(f_i(x_p, x), f_o(x_p, x))$ , that yields, as observed,  $f^*(x_i, \neg x)$  and  $f(x_p, x)$ .

### 5.2. On Undecidability

Without passing derivability problem, then, we will never face the problem of undecidability for that, applying meso-level properties as such, any theorem should not necessarily be proven because it is partly context-sensitive, and any theorem should not necessarily consistent with each other since there are some with context-sensitive and also some with context-insensitive characteristics.

Even if derivability problem is passed – i.e.: if mind is completely mechanical or automated by definition 1 – and undecidability is present, then for an external observer, there is *possibly* a chance that such undecidability is emerged by two different micro-level state in virtue of multiple realizability thesis.

**Lemma 2 Multiple Realizability** There are possibly  $f(x_i, x)$  and  $f(x_p, \neg x)$  and according to the observer, the inconsistency is due to the fact that  $f^*(f_i(x_p, x), f_o(x_p, x))$ , that yields, as observed,  $f(x_i, \neg x)$  and  $f(x_i, x)$ . But the derivation function  $f^*(f_i(x_p, x), f_o(x_p, x))$  is obtained from lower level in forms of  $g(\xi_i, \xi_p; x)$ ,  $g(\xi_i, \xi_o; x)$ ;  $g(\delta_i, \delta_p; x)$ ,  $g(\delta_i, \delta_o; x)$ . One might argue that the reduction to lower micro-level may lead to another inconsistency, but it entails more and more complete information about the universe and also reduction *ad infinitum* that are empirically implausible.<sup>5</sup>

Autonomous, learning, self-organizing mind, according to an external observer, as computation and as neurophysiological phenomena, will leave the shifting micro-level intact even though at the macro-level, provided that there is a Dedekindian infinity, an inconsistency occur.

Now let us look from the other way around. Imagine we are the observed with a prime observer existing external to us, probably a psychologist-physician mixture of a kind. Since we are only allowed to think and believe through propositional attitudes, then it is easy for ourselves to see our own undecidability.

**Lemma 3 Undecidability\*** Let  $f^*(x_p, x) \Leftrightarrow b(x_p, x) \downarrow$  such that  $b$  is another derivation function that says for each  $f^*(x_p, x)$ ,  $b$  can determine whether we will reach  $x_o$  from  $x_i$  (indicated by  $\downarrow$ ). If  $f$  is diagonalized such that  $f^*(x_p, x)$  then it follows that  $f^*(x_p, x) \Leftrightarrow b(x_p, x) \downarrow$ . As we continue, it is possible to have  $x^*$  such that  $f^*(x_p, x) \Leftrightarrow b(x^*, x) \uparrow$ . With another diagonalization, we can have  $f^*(x^*, x^*) \Leftrightarrow b(x^*, x^*) \uparrow$ . Then it is clear that we have both  $f^*(x^*, x^*) \Leftrightarrow b(x^*, x^*) \downarrow$  and  $\neg f^*(x^*, x^*) \Leftrightarrow b(x^*, x^*) \downarrow$ .

Following Lemma 3, it is easy to find ourselves trapped in undecidability as easy as finding antinomies Gödel (1931:175) himself once mentioned. But for a prime observer outside to us, this may be just another peculiarity of us having different micro-level states. And consequently, even if our computationally-analogous mind is undecidable, there is someone outside to us who observes that, for instance, we are still alive and immediately leave for other useful mental representation guided by some other micro-level state.

<sup>5</sup> Reductive physicalists and non-reductive functionalists' debates often assume complete information of the universe as if it were a closed system, while this counteracts the fact that limitation of our perceptual system only puts the universe to be an open system. A reductionist program – to gain emergent properties as necessitated – is bounded by our incomplete information about the universe, some essential relations are missing, while functional regularities are just too strong to ignore. Hence it is a matter of empirical difficulty Nagel points up such that this 'mild' reductionism corresponds to that of, e.g.: John Bickle's intertheoretic-reduction (1998).





If completely implemented in computer with a design able to make the computational mind autonomous, such micro-level state that emerges unpredictable macro-level state is possible to exist. Implemented computational mind hence becomes less automated, the propositional attitudes involved become less deductive to each other, and the formal axiomatic system becomes weaker. Dynamically speaking, the autonomous mind itself is *not* necessarily complete *or* consistent over time.

Conclusively, undecidability does not necessarily kill us and our autonomous machines or stop our cognitive minds and cognitive machines to function properly, at least – reductionistically – if fully-functioning computational mind is indicated by specific *pattern* of shifts in lower-level that is multiply realized. Undecidability does not kill us or stop our cognitive minds to function, at least – functionally – if functioning computational mind is indicated by us being able to perform propositional attitude manipulation such that anytime undecidability occurs, we can directly move to other theorems without bothering to solve such abstract logical concept. Undecidability kills us as an external observer, who for the rest of our lives, always tries to find a proof for an undecidable proposition in a setting such that Turing's halting problem occurs. But we, being perfectly normal human being, are not so stupid to find such proof until we die, and neither are our machines. Only a stupid (Lucasian) machine will do that.

We are no machine in this sense, but, machines that operate in many different levels. Hence, our micro-level says something more when we hear comical Gödel-type sentence such as 'This statement is unprovable' such that what is true is, unlike Lucas, not necessarily incomputable. Even further, it is not only a matter of completeness in finding a proof of a theorem in one level, but also completeness throughout inter-related levels although, if an autonomous design is implemented, such completeness probably will not be reached, but in advance, the machines may just start something else; it is out of control of our initial design.

The question is not about us being non-computable or not such that it is not or not precisely that computationalism is implausible – computational mind is just an analogy anyway. The question is about how to understand ourselves with the help of computer in many methods, many algorithms, many ways more than just manipulation of symbols in first-order or second-order logic.

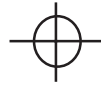
## 6. Acknowledgement

The author thanks Surya Research Int'l and Yohanes Surya for financial support, and all colleagues in Bandung Fe Institute for comments on preliminary version of this paper.

## 7. References

- Bates, E., & J. Elman. (1993). 'Connectionism and the Study of Change'. In M. Johnson., (ed.), *Brain Development and Cognition: A Reader*:623-42. Oxford: Blackwell.
- Bechtel, W. & J. Mundale. (1999). 'Multiple realizability revisited: Linking cognitive and neural states'. *Philosophy of Science*, 66:175-207.
- Bermúdez, J. (1995). 'Nonconceptual Content: From Perceptual Experience to Subpersonal Computational States'. *Mind and Language*, 10(4):333-69.
- Bickle, J. (1998). *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.
- Block, N. (1997). 'Anti-Reductionism Slaps Back'. *Philosophical Perspectives*, 11: *Mind, Causation, World*:107-33.
- Carruthers, P. (2003). 'Moderately massive modularity'. In A. O'Hear. (ed.), *Mind and Persons*. Cambridge: Cambridge University Press.
- Chalmers, D. (2002). 'The Components of Contents'. In D. Chalmers, *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.





- Feldman, J., & D. Ballard. (1982). 'Connectionist Models and Their Properties'. *Cognitive Science*, 6:205-54.
- Freeman, W. (2000). 'Brain Dynamics: Brain Chaos and Intentionality'. In E. Gordon., (ed.), *Integrative Neuroscience: Bringing Together Biological, Psychological and Clinical Models of the Human Brain*:163-71. Sydney: Harwood Academic Publishers.
- Fodor, J. (1998). 'The Trouble with Psychological Darwinism'. *London Review of Books*, 20(2).
- Fodor, J. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.
- Fodor, J., & Z. Pylyshyn. (1988). 'Connectionism and cognitive architecture: A critical analysis'. *Cognition*, 28(1-2):3-71.
- Gödel, K. (1931/1962). 'On formally undecidable propositions of Principia Mathematica and related systems 1'. *Monatshefte für Mathematik und Physik*, 38:173-98. Trans., B. Meltzer.
- Grush, R. & P. S. Churchland. (1995). 'Gaps in Penrose's Toilings'. *Journal of Consciousness Studies*, 2(1):10-29.
- Harnad, S. (1994) 'Computation Is Just Interpretable Symbol Manipulation: Cognition Isn't'. *Minds and Machines 4: Special Issue on "What Is Computation"*:379-390.
- Laurence, S., & E. Margolis. (2002). 'Radical concept nativism'. *Cognition*, 86: 25-55.
- Lucas, J. (1961). 'Minds, Machines and Gödel'. *Philosophy*, XXXVI:112-27.
- Lucas, J. (1998). 'The Implications of Gödel's Theorem'. *Talk given to the Sigma Club, London*.
- McCarthy, J., & P. Hayes. (1969). 'Some philosophical problems from the standpoint of Artificial Intelligence'. In B. Meltzer., & D. Michie. (eds.), *Machine Intelligence 4*. Amsterdam: Elsevier.
- Meehl, P., & W. Sellars. (1956). 'The Concept of Emergence'. In H. Feigl, & M. Scriven, (eds.), *Minnesota Studies in the Philosophy of Science, Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*:239-52. Minneapolis: University of Minnesota Press.
- Nagel, E. (1961). *The Structure of Science*. London: Routledge and Kegan Paul.
- Nagel, T. (1974). 'What is it like to be a bat?'. *Philosophical Review*, 83:435-50.
- Putnam, H. (1975). 'Philosophy and Our Mental Life'. In H. Putnam. *Mind, Language, and Reality*: 291-303. New York: Cambridge University Press.
- Searle, J. (1980). 'Minds, Brains, and Programs'. *Behavioral and Brain Sciences*, 3:417-57.
- Shapiro, S. (1993). 'Belief spaces as sets of propositions'. *Journal of Experimental and Theoretical Artificial Intelligence* 5(2-3):225-235.
- Shapiro, S. (1995). 'Computationalism'. *Minds and Machines*, 5(4):517-24.
- Sperry, R. (1991) 'In defense of mentalism and emergent interaction'. *Journal of Mind and Behavior*, 12:221-45.
- Suroso, R. (2004). 'Edukasi Natural dan Arsitektur Kognitif: Metode dan material edukasi dari perspektif sains kognitif'. *Working Paper WPR Bandung Fe Institute, July 2004*.
- Turing, A. (1950). 'Computing machinery and intelligence'. *Mind*, 59:433-60.
- van Gelder, T., & L. Niklasson. (1994). 'Classicalism and cognitive architecture'. In *Proceedings of the 16<sup>th</sup> Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

