Logique & Analyse **189–192** (2005), x–x

ON THE VERISIMILITUDE OF ARTIFICIAL INTELLIGENCE*

RODRIGO GONZÁLEZ AND ROGER VERGAUWEN

*This article originated as a paper for a meeting at the Information and Modality group at the Catholic University of Leuven. We should like to thank the organizers and participants of this meeting for helpful discussion. Many thanks also for valuable comments and critical evaluations to Frederick Truyen and Jan Heylen.

Abstract

Abstract. This paper investigates how the simulation of intelligence, an activity that has been considered the notional task of Artificial Intelligence, does not comprise its duplication. Briefly touching on the distinction between conceivability and possibility, and commenting on Ryan's approach to fiction in terms of the interplay between possible worlds and her principle of minimal departure, we specify verisimilitude in Artificial Intelligence as the accurate resemblance of intelligence by its simulation and, from this characterization, claim the metaphysical impossibility of duplicating intelligence, as neither verisimilarly nor convincingly simulating intelligence involves its duplication. To this end, we argue by a representative case of simulation that, albeit conceivable, Turing's test for machine intelligence wrongly equates the occurrence of indistinguishable intelligence performance to intelligence duplication, which is grounded in a *prima facie* conceivable but metaphysically impossible view that separates intelligence from its origin. Finally, we establish the following criterion for evaluating simulation in Artificial Intelligence: *simulations* succeed in AI if and only if they

The notional task of Artificial Intelligence is to simulate intelligence,

not to duplicate it.

Hilary Putnam in *Renewing Philosophy* (Putnam 1992, p. 11)

Introduction

Analytic studies of the nature of fiction have greatly multiplied in recent years. In addition to Meinong, Frege, Russell, the logical empiricists, Kripke, Putnam, Searle, Lewis and others, several approaches have broadly addressed this issue to date. Against this background, however, there has been little analysis of a common feature shared by fiction and Artificial Intelligence, namely, how both are able to carry out *verisimilar* simulations that comprise successful deceptions, which involve planned actions that cannot be

correctly explained in terms of verification in the state of affairs. While fiction epistemically devises imagined worlds that resemble the actual world by imitating its relevant properties, which makes the reader believe that a fictional text *is* the actual world, the mainstream of Artificial Intelligence assumes that carrying out good intelligence simulation by deception necessarily comprises intelligence duplication, as the Turing Test seeks to show. In this sense, this test, which establishes when intelligence can be attributed to computers and robots, conceives that a computer behaving *as though* it understood questions when replying meaningful answers is *sufficient* for regarding it as intelligent (Turing 1950). Tacitly considering that successful intelligence simulation can be equated with its duplication, such a stipulation has encouraged the majority of Artificial Intelligence researchers to believe that the artificial duplication of mental life by its simulation is not only theoretically plausible, but also technically feasible in the long run.

Taking into account how the Turing Test attributes intelligence to machines, this article will explore how verisimilitude occurs both in fiction and Artificial Intelligence, and explain why from intelligence simulation it does not follow intelligence duplication. To avoid dealing with the fuzzy notion of understanding and the variety of replies to Searle's Chinese Room argument (Searle 1980 and 1990), this investigation will particularly focus on the implausibility of creating intelligence by pure verisimilar and, yet, successful simulations, which will further discuss Searle's insight that simulation involves no duplication in AI. Touching on the distinction between imagining and conceiving, section 1 characterizes how conceivability and possibility are connected with possible worlds. Commenting on Ryan's account of fiction as a departure from the actual world ruled by the principle of minimal departure, section 2 addresses verisimilar fiction in terms of conceivable worlds, and explains how they are connected with possible worlds by the agent's ability to epistemically imagine worlds different from the actual

world. Section 3 examines how AI's successful simulations, which abide by the principle of minimal departure, do not entail the duplication, but the *resemblance* of human intelligence. Finally, section 4 criticizes the equation between intelligence simulation and duplication by a plausible representative case of simulation, the Stradivarius violins' case, which illustrates that, since *verisimilar* simulations do not require the duplication of what is being simulated, Artificial Intelligence need not duplicate the human intellectual capabilities to accomplish its task. From this argument, the article advances a criterion for good simulation in Artificial Intelligence, according to which AI's verisimilar simulations sufficiently imitate human intelligence if and only if they are able to persuade people that intelligence has been duplicated, which requires that intelligence simulations minimally depart from actual intelligence. The article poses a final problem for any functional reduction of intentionality sought by GOFAI ('Good Old Fashioned Artificial

Intelligence'), showing again why simulation cannot be equated with duplication from a theoretical standpoint.

1. *Conceivability, possibility and the connection between fiction and reality*

Schlick's early work "Positivism and Realism" attempts to shed light on how meaningful propositions are connected with reality — or the *given*. On his view (Schlick 1932/1933, p. 42), "what is merely empirically impossible still remains thinkable; but what is logically impossible is contradictory, and cannot be therefore thought at all." While propositions are thinkable when their verification is plausible, bogus propositions convey no meaning, as they reveal unverifiable but conceivable situations. For example, '*zeutron*', an imaginary atomic particle whose existence could not be assessed by any possible empirical means, yields meaningless sentences, for, on Schlick's view,

it is impossible to state what the world would be like if the proposition 'X *is zeutron*' were true or false.

But, if uttered, what psychological processes would be involved if one posited this chimerical entity, the existence of which is conceivable and, still, unverifiable? Would one imagine or conceive 'X *is zeutron*'? Would there be any important difference between conjuring up a '*zeutron*' and supposing its existence? These questions suggest that a brief characterization of the difference between imagining and conceiving is necessary to understand how it is possible to utter sentences like 'x *is zeutron*', which are regarded as absurd and meaningless in relation to the states of affairs. Despite the fact that both imagining p and conceiving p can be regarded as *intentional mental actions* (Mele 1996, p. 233) which evince representational content with direction of fit, two different attitudes apparently arise from imagining p and conceiving p.

Traditionally, imagining p has been associated with thinking of an event that is not currently present to the senses, which necessitates the concurrence of percepts and memories. According to Yablo, imagining p *requires* an agent as a measuring device, who posits p, keeps track of the salient features of the actual world, and uses the perceptual faculty of imagination (one's mind eye) to mentally simulate looking at an object or event in which p (Yablo 2002, p. 457–58). Such a mental simulation supposes a centered world in which agents envisage themselves being struck by imagined objects or events. By combining past experiences, one is able to simulate the activity of looking at p by conjuring up mental imagery which, despite creating a fantastic or a disturbing entity or event, allows the *internal* verification of p. This explains why it is so difficult for people born blind to imagine colors and, hence, blue apples, flying pigs, red dogs, or any manipulated

image arisen from the perceived states of affairs of the actual world[1]. To fancy a flock of flying pigs, for instance, one has to manipulate memories of pigs, past experiences of birds or flying objects to encourage the imagination to create an image of pigs flying in V-shape formation, for instance, which would indeed verify 'pigs fly.' Thus, the perceptual imagination recombines images to envisage chimerical entities, the existence of which may count as utterly implausible in the state of affairs. Still, these images can be evaluated from an agent's viewpoint, which inspired David Hume to hold that, eventually, everything, even impossibilities, can potentially be *imagined*:

[1] Indeed it remains controversial whether or not people born blind may successfully conjure up images to imagine p. If so, the question to ask would be: Are those images *correct* as to what they are supposed to represent? People born blind might analogically imagine red. Nevertheless, by doing so, they would never imagine what a person with the sense of vision does, because the former would only use analogical past experiences to imagine red.

> "Tis an establish'd maxim in metaphysics, *That whatever the mind clearly conceives includes the idea of possible existence,* or in other words, *that nothing we imagine is absolutely impossible*. We can form the idea of a golden mountain, and from thence conclude that such a mountain may actually exist." (Hume 1978, p. 32, italics in original).

Nevertheless, as described above, there is another quite compelling aspect of imagining p, namely, the ability of an agent to conceive, suppose or entertain a statement S without images, which arises from the agents' capacity of adopting a purely epistemic stance towards p. As conceivability is a property of statements, it is perfectly possible that an agent imagine situations — by conceiving or supposing them — which might have not been caused by previous experiences or perceptual images. This imagining is not directly

grounded in imagery. For example, conceiving an atom of lead, undetectable colors, molecules of H_2O, or Germany winning the Second World War does not necessarily require the participation of perceptual images. In these particular cases, conceiving p aims at having *an intuition of a world* W, in which p would be true or false if evaluated in W. In those cases, one *imagines* that p in the second sense of imagining, that is, in the sense of epistemically supposing a situation, which indeed helps explain why although 'X *is zeutron*' is not verifiable in principle, and it lacks meaning on Schlick's view, one is able to *imagine* a situation in a world W, where 'X *is zeutron*' might *epistemically* hold true. Note that, even so, it is not possible to think of states of affairs in the actual world that yield 'x *is zeutron*' true, because, if uttered, one would not be able to evaluate what the actual world would be like in case 'x *is zeutron*' was true or false. From a scientific standpoint, nevertheless, the latter is the only relevant *semantic evaluation* procedure, as analyzing the truth conditions of p constraints one to find the cognitive strategies that

are most likely to lead to the set of empirical possible consequences of p in the states of affairs of the actual world. This last aspect is indeed the kernel of the Schlick's criterion, although he never thought of the widely discussed distinction between the actual and possible worlds.

In this respect, yet helpful in explaining what the empirical evaluation of scientific propositions is like, Schlick's semantic criterion of verifiability in principle seems ineffective in other dominions, especially in view of the distinction between imagining, conceiving and the widely held modal view of possible worlds. Further, the application of Schlick's criterion in fields that include statements describing understandable, but still unverifiable statements in the state of affairs of the actual world shows that such a criterion seems to lack the proper semantic *finesse* at this level. In particular, the analysis of fiction in terms of verifiability in principle is difficult and lacks depth here, as fiction creates events that are imaginable, credible, communicable and generally verisimilar. Invented stories create scenarios

in which characters 'come to life' and communicate their thoughts through plots, which interestingly make stories become *intelligible, believable and plausible*. However, unlike scientific statements which are verifiable *in principle*, most fictional narrations are regarded as widely *conceivable*, despite not entailing verifiable conditions in the state of affairs of the actual world.

The agents' epistemic abilities to imagine verisimilar fictional worlds and situations suggest that another approach is indeed required to account for the fictional gesture. This feeling intensifies as soon as one seriously considers the issue of whether computers and robots are intelligent, as Artificial Intelligence has vigorously claimed. Is the aim of Artificial Intelligence imaginable, conceivable or plausible, given the conditions from which intelligence arises in the actual world? An analysis of how fiction makes stories intelligible and believable, via the concept of verisimilitude, will indeed help answer this question.

2. *Fictional verisimilitude via the principle of minimal departure*

Despite being hardly verifiable in the states of affairs of the actual world, the verisimilar character of fictional statements raises the following questions: Do verisimilar fictional statements convey any possible connection with the state of affairs of possible worlds? If so, do they yield truth or falsehood, as any other semantically meaningful proposition? And, are fictional characters and plots somehow *real* from a metaphysical standpoint? As fiction narrates *conceivable* situations appealing to the imagination — perceptual or not — and/or to ways the world could have been different, the modal notion of possible worlds seems to be the most suitable method to analyze its nature. Ryan adopts this approach, explaining the fictional gesture in terms of the framework of modal logic and the semantics of possible worlds, and proposing a theory to analyze fiction by means of juxtaposing and connecting fictional worlds with possible worlds (Ryan 1991). There are, however,

two elements worth bearing in mind when examining the relation between fiction and possible worlds.

On the one hand, the nature of possible worlds is still philosophically controversial. Although the actual world is admittedly the source from which possible worlds are accessed via counterfactual reasoning, the status of possible worlds is yet debatable. At first glance, possible worlds are as real as the actual world, because, by intuition, every possible world could have been the actual world (Lewis 1979, pp. 182–184). However, on close analysis, it turns out that statements about possible worlds, unlike propositions about the actual world, are mentally dependent on how things actually are, but not vice versa (Rescher 1979, p. 169). Undoubtedly, this suggests that possible worlds are mentally dependent. Whether or not alternative possible worlds — or APWs — are as *real* as the actual world is assessable by supposing the following situation: a world in which Hitler had won the war (Ryan 1991, p. 19). Although this second Hitler might eventually imagine a third Hitler

losing the war, the latter Hitler would not have the same ontological status as the actual one, for the third Hitler would owe his existence to the *recursivity*, rather than to the *reversibility*, of the relation of *alternativeness*. That is, provided one travelled to an APW where Hitler had won the war, and from there to one of its *alternatives* where he loses the war, one would be unable *to return* to the actual world where Hitler actually lost the war. This naturally follows from the fact that the third imagined Hitler has been created by recursion, i.e., by postulating a new possible world from a possible world. Consequently, this argument shows that possible worlds depend upon the actual world as well as upon the agents' ability to postulate these worlds.

On the other hand, a large number of philosophers have traditionally held that *any possible statement is conceivable but not vice versa*. As Putnam succinctly puts it:

> "We can perfectly well imagine having experiences that would convince us (and that would make it rational to believe that) water isn't H_2O. In that sense, it is conceivable that water isn't H_2O. It is conceivable but it isn't logically possible! *Conceivability is no proof of logical possibility*" (Putnam 1975, p. 233, our italics)

This remark apparently flies in the face of Ryan's account of fiction, as there are a number of fictional worlds which are indeed conceivable, but may not be regarded as possible worlds, for they are not entirely maximally logically consistent, which is a necessary condition of possible worlds. Furthermore,

possible worlds have traditionally been considered as maximally consistent sets of propositions (Copeland 2002, p. 104)[2]. Therefore, there would be a number of fictional worlds which could not be analyzed in terms of possible worlds, unless one gives up classical logic.

However, there is no need to appeal to paraconsistent logic to modally account for fiction. Putnam's particular case of conceiving the statement 'water isn't H_2O' appeals to the second meaning of imagination touched in section 1, that is, to supposing a situation without mental imagery. Indeed, when one envisages the possibility that water isn't H_2O, one mentally simulates

[2] The idea of possible worlds as maximally consistent sets of propositions can be traced back to the very notion of *state-description*, which is 'a class of sentences that represents a possible specific state of affairs by giving a complete description of the universe of individuals with respect to all properties and designated by predicates in the system... A state-description contains for every atomic sentence S_i either S_i itself or $-S_i$, but not both...' (Carnap 1946, p. 50) Moreover, Carnap characterized the state-descriptions as representing Leibniz' idea of possible worlds or Wittgenstein's idea of possible states of affairs (Carnap 1947, p. 9).

a situation by reflecting on what an *epistemically imagined world would be like* in case water wasn't H_2O, e.g. one considers how the world might be, provided that God had created it with different natural laws. In this sense, imagining a world in which 'water isn't H_2O' is perfectly tenable from an epistemic viewpoint, but metaphysically impossible, for the identity between water and H_2O is metaphysically necessary (Kripke 1980). Even so, to *imagine* that God has created a world W with different laws, it is not necessary to conjure up perceptual images, but an *incomplete* conceptual configuration of W, leaving unspecified the irrelevant elements for entertaining 'water isn't H_2O'. Thus, when agents *imagine* or suppose 'water isn't H_2O', they perform an intentional mental act, which configures an imaginable world W where one simply *conceives, supposes, or mentally simulates* incomplete apparent events and situations in a world W, all of which may *internally* verify 'water isn't H_2O'.

Most importantly, the intentional act involved in entertaining such a proposition does not entail the existence of the *apparent event, situation or world*. This explains why conceiving metaphysical impossible statements like 'water isn't H_2O', 'Hespherus is not Phosphorus', 'Karol Wojtyla is not John Paul II', or even contradictory statements is plausible from the agent's *epistemic* viewpoint, but not from the point of view of metaphysical necessity. On close examination, all these imaginings are metaphysically impossible, which does not *prima facie* prevent conceiving false *a posteriori* necessary identities as *mere imaginings*[3].

[3] We wish we had the opportunity of getting deeper into the discussion of whether conceivability is a guide to possibility (Chalmers 2002), especially in relation to what several philosophers have claimed about 'water isn't H_2O' as conceivable and yet metaphysically impossible. Unfortunately, a closer examination of this issue would lead us far away from discussing the verisimilitude of Artificial Intelligence.

This rough distinction of conceivability and possibility also allows explaining the creation of fictional entities and worlds and, thus, the semantic plausible evaluation of fictional statements. Whenever there is a metaphysically impossible conceived fictional situation, the agent can evaluate its truth-value solely from an epistemic stance, as in 'Let's pretend that Karol Wojtyla hadn't been John Paul II.' This clarification provides the basis for understanding why fictional worlds do not seem fully compatible with possible worlds, as the latter must evince maximal consistency. Unlike possible worlds, inconsistent fictional worlds can be imagined, conceived or mentally simulated, portraying situations which may or may not include perceptual imaginings, and moreover, leaving several details unspecified in a fictional world W. Additionally, it is quite possible to account for fictional contradictory statements, and inconsistent fictional worlds by envisaging or assessing these statements from a purely epistemic viewpoint. Precisely, this last approach captures the quintessence of fiction, namely, its ability to narrate

stories which enthral the reader by appraising counterfactual situations, *and even ones that, albeit conceivable, are impossible from a metaphysical viewpoint*. A satisfactory account of fiction must naturally explain how fiction makes conceivable the impossible.

In this regard, the following proposed Principle of Conceivability for Fictional Worlds helps account for contradictions and inconsistencies in non-fully consistent *purely* conceivable fictional worlds:

The Principle of Conceivability for Fictional Worlds: As any imaginable fictional situation is conceivable, conceivable fictional worlds and situations, which are *prima facie* believable and intelligible but metaphysically or textually impossible on close examination, may include contradictory and inconsistent statements.

Note that, rather than being incompatible with Ryan's account of fiction, this principle explains why, while possible worlds exclude contradictions and inconsistencies, some worlds of fiction may not, e.g. those that belong to the theater of absurd, poems, postmodern tales and so on so forth.

Despite not considering the particular connection between conceivable worlds, fictional worlds and possible worlds, Ryan's modal account of fiction allows one to investigate how the reader engages in fictional worlds by the concept of *recentering*, or how fictional worlds comprise a textual universe at the center of which lays the 'textual actual world' (TAW) (Ryan 1991, pp. 22–24). This new actual world is an external representation or the Textual Represented World (TRW), which consistently *makes the reader believe that* TAW *is* AW.

Incidentally, Ryan's following approach helps explain how fictional stories illustrate situations that portray *a new conceivable/possible world*:

> "Since we regard 'the real world' as the real of the ordinary, any departure from norms not explicitly stated in the text is to be regarded as a gratuitous increase of the distance between the textual universe and our own system of reality. ...
>
> ... We can derive a law of primary importance for the phenomenology of reading. This law — to which I shall refer as the Principle of Minimal Departure — states that we reconstrue the central world of a textual universe in the same way we reconstrue the alternate possible worlds of non factual statements: as conforming as far as possible to our representation of AW." (Ryan 1991, p. 51)

The principle of minimal departure allows characterizing how deceptions, forgeries, and deceits successfully operate for the *verisimilitude* they exhibit

when mimicking situations that effectively deceive people. If not, those actions do not effectively persuade people. Although Ryan does not appeal to the concept of verisimilitude itself, a concept such as this one is crucial to understand how successful deceptions, which minimally depart from the actual world, operate efficiently convincing people. In fiction, deception stems from the dynamics between narrator, plot, and reader, which makes people believe that TAW *is* AW if and only if the former verisimilarly *imitates* the latter in relevant respects.

While gripping verisimilar fictional stories become intelligible and believable for readers, inconsistent fictional stories tend to discourage the reader to believe them, as inconsistent events admittedly affect the plausibility of a story. A necessary element of a successful deception D is its intelligible, believable, plausible character, which will naturally deceive the readers by making them believe a story. For example, suppose a fictional story providing a setting with extreme conditions under which Jones wasn't able to

quench his thirst, even if the author had not expressly suppressed the ability of satisfying thirst for water. Provided that Jones found water and drank a good deal of it, the story would make no sense whatsoever, despite narrating a conceivable situation. Still, a story like Jones', which makes no sense due to its slip-up, would hardly be believable and appealing for the readers. On the contrary, if a different story described that Jones found H3O water in Mars, quenched his thirst after drinking a good deal of this particular water, and reported to Earth that his age spots started disappearing due to some yet unknown healthy properties associated with H3O, this new story would indeed make sense. The fictional situation described simulates what might plausibly occur in the actual world, given the conditions specified by the story and the properties that the fictional world imitates of the actual world. As the second story illustrates, fictional situations not only simulate the relevant properties of the actual world, but also leave unspecified all the details which are irrelevant for making the story believable.

Fiction plays a quite specific role in *mentally simulating* that a fictional world *is* an actual world by its proper imitation. Imagined worlds depicted by fictional stories become enthrallingly believable if and only if they meet this very specific condition: *verisimilitude*, which is the ability of a story to *appear real* by emulating specific properties or events in the actual world. Describing them as necessary for abiding stories by PMD and, thereby, making them verisimilar, Ryan lists the following relevant identity relations between a textual actual world and the actual world (Ryan 1991, pp. 32–33): identity properties, identity inventory, compatibility inventory, chronological compatibility, physical compatibility, taxonomic compatibility, logical compatibility, analytical compatibility, and linguistic compatibility.

The interplay between these identity relations creates different verisimilar fictional worlds, helping to provide the setting, and determining what degree of credibility the story will hold. As a result of abiding by PMD, a story will become *verisimilar* if it conceives a world that accurately resembles the

actual world. If not, then:

1. The story will partially lack verisimilitude
2. Yet conceivable, the story will hardly be intelligible
3. The reader will regard the story as hardly credible

All things considered, Ryan's approach to fiction in terms of possible worlds as well as the clarification of the difference between conceivability and possibility, allows the characterization of how verisimilitude arises, that is, from fictional texts that minimally depart from the actual world. But, most importantly, this analysis, which helps characterize fiction as the process that generally creates verisimilar fictional worlds, can be extrapolated to another field in which verisimilitude plays a crucial role, namely, Artificial Intelligence. Devising a method to replace the question 'Can machines think?', the Turing Test (Turing 1950) purportedly provides a method to assess whether

digital computers *are* intelligent. Adequately programmed to simulate human beings, the computer deceives human interrogators and passes the test. This method abides by the principle of minimal departure, because the successful processes of simulation carried out by Artificial Intelligence evince high degrees of *verisimilitude*, which allegedly provides sufficient evidence to conclude that computers or robots *duplicate* human intelligence.

3. *AI's verisimilar simulations as minimal departures from human intelligence*

In the previous section, the examination of the notion of verisimilitude and the quite specific role it plays in fiction showed that the simulation of specific properties of the actual world in a textual actual world deceives people when

reading books. Verisimilar mental fictional simulations assume that the textual actual world *is* the actual world. In this section, we will extrapolate the notion of verisimilitude to the context of Artificial Intelligence so as to assess whether or not Turing's view on the aim of this discipline, which takes for granted that intelligence simulation comprises the duplication of human intelligence, is achievable *in principle*. To accomplish this task, a brief examination of the historical context in which AI was born will not only serve to grasp its aim, but also to explain why a large number of AI researchers have equated human intelligence *simulation* to human intellectual abilities *replication*, ever since the Turing test was conceived in the 50's.

Although it is widely admitted that Alan Turing paved the way for the foundation of this discipline, as it is known today, the exact birth of Artificial Intelligence as such is hard to pin down. In fact, it has been held that the basis for Artificial Intelligence had somehow been provided under the name of *machine intelligence* in Britain earlier in the 40's (Copeland

2000, p. 1). As this logician explicitly suggests, a number of AI researchers, including Turing himself, had thought of designing machine learning and heuristic problem-solving programs, and even machines able to play checkers or chess. In particular, Turing had already thought of the idea of machine intelligence by 1941 and chess playing machines by 1945.

Is it possible to identify the exact event that gave rise to the project of Artificial Intelligence as such, given the growing enthusiasm with the idea of machine intelligence back then? Copeland regards the birth of Artificial Intelligence approximately in 1956,

> "the year in which a program written by Newell, Simon, and Shaw — later named the Logic Theorist — successfully proved theorems from Whitehead and Russell's *Principia Mathematica*, and also the year of John McCarthy's *Dartmouth Summer Research Project on*

> *Artificial Intelligence*, the conference which gave the emerging field its name" (Copeland 2001, p. 1).

Although those events may indeed count as the foundation of Artificial Intelligence, as it is known nowadays, Alan Turing had conceived how machine intelligence ought to be carried out before, around 1948. That year Turing had already advanced an idea that would have a decisive influence on the Turing test as well as on AI's researchers, and how they ought to create and assess intelligence. In the context of developing a paper chess machine, Turing provided the basis for Artificial Intelligence's aim by describing this *early version* of the Turing test:

> "It is possible to do a little experiment...even at the present stage of knowledge. It is not difficult to devise a paper machine which

> will play a not very bad game of chess. Now get three men as subjects for the experiment, A, B, and C. A and C are to be rather poor chess players. B is the operator who works the paper machine. (In order that he should be able to work it fairly fast it is advisable that he be both mathematician and chess player.) Two rooms are used with some arrangements for communicating moves, and a game is played between C and either A or the paper machine. C may find it quite difficult to tell which he is playing. (This is a rather idealized form of an experiment I have actually done.)" (Turing 1948, p. 23)

This chess game represents the *very* cornerstone of Artificial Intelligence, providing the basis upon which Turing devised the Turing Test as an experimental method to *test* whether or not machines are intelligent (Turing 1950). The Turing Test devises a game based upon a similar dynamic. However,

unlike the chess game, the imitation game in the Turing test consists of a scenario in which there is a programmed computer answering the questions is being put to by human interrogators. As the figure illustrates below, in the simplified standard version of the imitation game[4], there is a person inside one room (A), and a computer inside a second room (B). Outside these two rooms, there is a human interrogator (C), who passes typewritten questions to rooms A and B. This interrogator gets typewritten answers back, and has to decide who the person is by the answers.

Note that, by assessing the ability of the machine to answer questions as though it was a person, the Turing test is said to provide *sufficient evidence*

[4] For the sake of simplicity, we leave aside Turing's two former versions of the imitation game, which mainly base the deception process upon the inability of the interrogator to decide the gender of participants, and the replacement of a person for a programmed computer. In any case, the former versions of the imitation game stress the same point, that is, the simulation of a person's linguistic performance.

for concluding that the computer *is* intelligent. Sharply distinguishing between *the physical abilities* and *the intellectual abilities* of human being, the test assumes that the machine is able to simulate the latter by imitating a person's linguistic performance, which need not duplicate the brain itself.

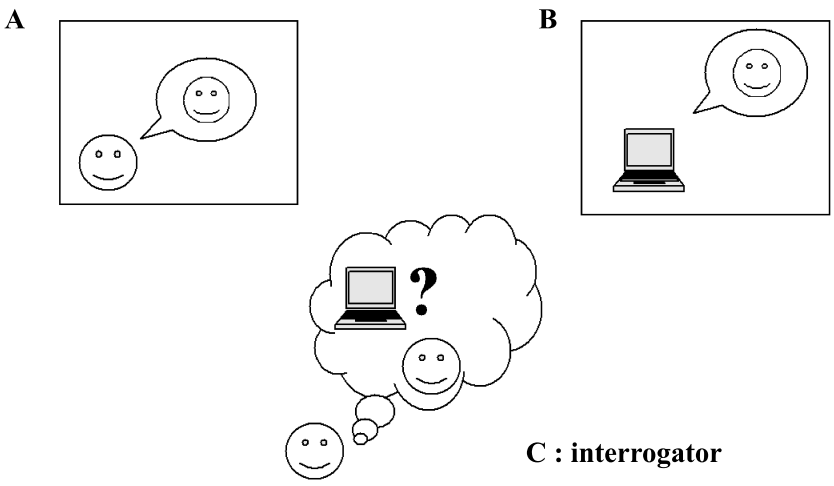

Figure 1

According to the test, if the correctly-programmed computer replies the correct typewritten answers to the questions posed by external interrogator and, thus, linguistically behaves like any person, it will deceive a majority

of interrogators, making them believe that there is a person inside room B. After deceiving 7 out of 10 or more interrogators, the computer is said to have successfully passed the test. The kernel of the Turing Test, accordingly, is the *deception* of a majority of interrogators, who will mistakenly believe there is a person inside room B, provided the computer's ability to display behavior that successfully imitates a person's intellectual capacities.

Saygin et al. explicitly claim the following in relation to the *nature* of the game:

> "Here is our explanation of Turing's design: The crucial point seems to be that the notion of *imitation* figures more prominently in Turing's paper than is commonly acknowledged. For one thing the game is inherently about deception. ... Alternatively, the TT for

> machine intelligence can be reinterpreted as a test to assess a machine's ability to pass for a human being." (Saygin et al. 2000, pp. 26–27, italics in original)

By focusing on *deception*, the Turing test allegedly provides compelling evidence to conclude that computing machines *are* intelligent. The game is supposed to be a well-defined *method* to determine *in principle* whether or not computers are intelligent, so long as they are able to *imitate* the human intellectual abilities. The test does not incidentally provide a *behaviorist or operational definition of intelligence*, which is what a number of philosophers and commentators of the test have wrongly claimed (Block 1990 p. 378, Hodges 1992 p. 415, and French 2000, p. 115).

Neither does the Turing test provide a sufficient condition of intelligence, as it has also wrongly been claimed. Strictly speaking, the computer's simulation of a person's linguistic performance *justifies* one to believe that the

duplication of human intellectual capabilities by simulation, despite the fact that man and machine may carry out something utterly different when answering questions about stories. After denying the importance of investigating further *the nature of the imitation* game, Turing considers such a possibility as follows:

> "May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, *a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection*. It might be urged that when playing the "imitation game" the best strategy for the machine may possibly be something other than imitation of the behaviour of a man. This may be, but I think it is unlikely that there is any great effect of this kind. In any case there

> is no intention to *investigate here the theory of the game*, and it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man." (Turing 1950, p. 42 our italics)

Oddly enough, the force and weakness of the Turing test precisely lies in the nature of the imitation game, for Turing considers passing the test as *sufficient evidence* for thinking that a computer has mind. Broadcasted by BBC in May 1951, Turing's lecture "Can a Digital Computer Think?" strongly emphasizes the same point as follows:

> I believe that [digital computers] could be used in such a manner that *they could appropriately be described as brains*. ... This ... statement needs some explanation. ... In order to arrange for our computer to imitate a given machine it is only necessary to programme the computer to calculate *what the machine in question*

> *would do under given circumstances* . . . If now some particular machine can be described as a brain *we have only to programme our digital computer to imitate it and it will also be a brain*. If it is accepted that real brains, as found in animals, and in particular in men, are a sort of machine it will follow that our digital computer suitably programmed will behave like *a brain*.
>
> . . .
>
> [O]ur main problem [is] how to programme a machine to imitate the brain, or as we might say more briefly, if less accurately, *to think*.
>
> (Copeland 2001, p. 11, our italics)

Undoubtedly, this confusion between describing a programmed machine as a brain with a brain has encouraged a large number of Artificial Intelligence

researchers to regard the aim of their discipline as the *duplication* of human intelligence. Indeed, Turing's sharp distinction between the intellectual abilities and the physical abilities of man explains the aforementioned confusion. Provided that this distinction was wrong, it would not be possible to programme a computer to imitate thought and, thereby, be a brain.

Turing's ambitious view on AI's goal has misled a large number of people, who claim that AI is committed to artificially creating intelligence by computational simulations of human intellectual capabilities, which supposedly comprises the duplication of intelligence. Among the large number of Artificial Intelligence definitions, this one stands out as the clearest in exhibiting the long-range substantial interest of AI's researchers:

> "Artificial Intelligence (which I'll refer to hereafter by its nickname, "AI") is the subfield of Computer Science devoted to developing

> programs that enable computers to display behavior that can (broadly) be characterized as intelligent. Most research in AI is devoted to fairly narrow applications, such as planning or speech-to-speech translation in limited, well defined task domains. But substantial interest remains in the long-range goal of building generally intelligent, autonomous agents." (Thomasson 2003)

The wide spread confusion between simulating and duplicating intelligence in AI raises these questions then: Is one really entitled to believe that the computer duplicates the human intellectual abilities if it successfully imitates them? And, does Artificial Intelligence need the *duplication* of intelligence to properly simulate it? Is the nature of the imitation game not relevant, as to the possible difference between simulation and replication? To answer these questions, it is quite useful to compare *the imitation* game in the Turing test

with the *verisimilar* make-believe process of fiction and in what respects they overlap and differ.

On the one hand, strictly speaking, neither fiction nor Artificial Intelligence requires the *duplication* of any property to verisimilarly make one believe that P. On the contrary, by *simulating* properties, both disciplines prompt imitation processes which suffice for making one believe that P. While the reader of a fictional text commits to the verisimilar representation of the textual actual world, the computer is said to simulate intelligence, making the interrogators believe it is a person. However, this deception brings about '*as if*' intelligent behavior, which does not provide *incontrovertible evidence* for thinking that the computer has *duplicated* the very same human intellectual capabilities of man. Hence, Artificial Intelligence need not duplicate Intelligence, but make the interrogators believe so by its verisimilar simulation.

On the other hand, both the Turing Test and fiction convince one that P, in spite of the fact that the answers and the narrated stories may lack veracity. However, there is a major difference between the reader's and the interrogator's attitude as to this point. Whereas the former willingly *commits* to believing that the textual actual world is the actual world, the computer simulating human intellectual abilities deceives the latter, who unknowingly believes there is a person instead of a programmed computer. And, unlike the reader accepting conventions to believe that the textual actual world *is* the actual world (e.g. 'Once upon a time there was...'), the interrogator does not engage in believing that the computer is intelligent intentionally. On the contrary, the interrogators simply participate in the imitation game, and get convinced that the computer is another person locked in a room.

In light of the comparison between fiction and Artificial Intelligence the important question to ask is the following: *Does the verisimilar simulation of intelligence actually entail its duplication?* The deception carried out by

the computer in the imitation game suggests that mimicking the linguistic performance of a person does not *entail any duplication*, because, *prima facie*, the successful verisimilar simulation of intelligence requires no intelligence duplication. Although intelligence simulation suffices for convincing interrogators that the digital computer *is* able to linguistically perform as any person and, thus, for regarding the machine as intelligent, the accurate verisimilar pretense does not entitle one to conclude that the computer has a mind. In this sense, Turing hastily takes for granted that the simulation of intelligence entails its duplication. From the viewpoint of the distinction between conceivability and possibility sketched in section 1, two additional interesting questions to ask are: Is the computer's human intelligence simulation identical to human intelligence duplication? And, is this alleged identity necessary or purely contingent from a metaphysical viewpoint?

4. *Why does AI's verisimilar simulation not entail* duplicated *performance?*

The Stradivarius violins' case and the Philosophy of Mind

A brief look at the role the notion of *origin* plays in defining identity seems quite convenient to tackle the issue of whether it is metaphysically possible to duplicate intelligence by simulating it. In the context of defining identity, Kripke asserts the following:

> "In the case of this table, we may not know what block of wood the table came from. Now could *this table* have been made from a completely *different* block of wood, or even of water cleverly hardened into ice — water taken from the Thames river? We could conceivably discover that, contrary to what we know think, this table is indeed made of ice from the river. But let us suppose that it is not.

> Then, though we can imagine making a table out of another block of wood or even from ice, identical in appearance with this one, and though we could have put it in this very position in the room, it seems to me that this is *not* to imagine *this* table as made of wood or ice, but rather it is to imagine another table, *resembling* this one in all external details, made of another block of wood, or even of ice."
>
> (Kripke 1980, pp. 113–114, italics in original)

This passage offers a revealing insight as to how two objects differing from origin, yet different, may look quite similar, which certainly has a connection with Artificial Intelligence and how this discipline carries out verisimilar simulations. *Origin* and its connection with metaphysical identity explains why simulation does not entail duplication, because if B has a different origin from A, then B will resemble A at the most and, yet, it will never be

identical to A. Likewise, if two objects share one property that is crucial to define them, they at least must have the very same property in every possible world. In other words, they necessarily have to share at least this defining relevant property. For example, if Socrates and Plato are human beings, they must have 46 chromosomes. Have two objects the same origin, they will necessarily share one respect or property that help define their identity.

This grasp of the notion of origin has also a close tie with the issue of authenticity. Consider the following example of a Stradivarius violins' forgery. Despite all the efforts of science and music experts, it is still a mystery what exactly causes the outstanding performance of these Cremonese violins of the late 17^{th} to 18^{th} centuries (Gough 2000). There are a number of hypotheses which attempt to explain their superior sound such as the varnish, the wood, a fungus that enhanced the quality of the wood, a little ice age that made the wood denser, and so on. All these hypotheses intend to pin down the origin and causal factors involved in their performance so that scientists

might duplicate them in the future. Suppose that two thieves stole the Stradivarius original handbook for manufacturing violins[5], and attempted to copy them in all respects by following the strict guidelines provided by the book. Imagine that a very special kind of fungus, which used to enhance the quality of the wood with which Stradivari crafted his violins and originated the sound of them, had completely disappeared nowadays. As the thieves could not employ the very same materials with which Stradivari manufactured his

[5] This example deals with the problem of whether or not simulation comprises duplication in the same vein as Cleland's objection to explaining mental phenomena by effective procedures (Cleland 1993). Whereas she argues that Turing Machine procedures involve neither causality nor mental life, our argument states that verisimilar simulations cannot duplicate the simulated properties or events, as machine intelligence and human intelligence arise from different *origins and prompt two different causal processes*.

violins, they would be unable to duplicate the remarkable sound of a Stradivarius violin. Even so, the crooks would be able to rip off several people, crafting excellent replicas.

This case raises two important issues. On the one hand, whether or not forgery or simulation can be sufficient for duplicating an original authentic thing, in this case, a Stradivarius violin. On the other hand, whether or not one could *possibly* duplicate the Stradivarius violins' performance by adequately simulating them, as Artificial Intelligence might claim, for instance. Consider another possibility associated with the Stradivarius violins, which will directly deal with AI's equation of simulation to duplication.

On the face of the Stradivarius violins' mystery, could a computer program duplicate the Stradivarius violins' performance by verisimilarly simulating it? Suppose that AI researchers developed a program to resolve the mystery of Stradivarius violins, which additionally could verisimilarly simulate their performance. To this end, they design a program called MESSIAH, in honor

to the most celebrated violin in the world crafted by Stradivari in 1716: 'The messiah.' Imagine, further, that MESSIAH was able to successfully pass a test analogous to the Turing test, that is, by deceiving judges who could not distinguish the computer's simulation from that of an authentic Stradivarius violin's performance.

This 'Stradivarius test' would run as follows: a violinist plays on a Stradivarius violin a limited set of compositions at the request of some judges. All those compositions are magnificently replayed by MESSIAH, which is run on a computer equipped with a high fidelity sound system. The judges have to assess the performance of the Stradivarius violin and the computer through headphones, and have to decide which one the Stradivarius violin is. MESSIAH is programmed to mimic the very same human-like sounds a violinist player makes when playing violin, and even to make extremely rare faults, which prevent favoring the violinist. After several compositions were played, most judges would be unable to tell the difference, and decide

which performance belongs to the Stradivarius violin, because the computer would perform in a manner undistinguishable from that of the violin, deceiving the majority of them. By doing so, MESSIAH would indeed pass the Stradivarius Test.

Being able to accurately imitate performances, computer programs such as MESSIAH and its verisimilar simulation cannot duplicate those performances. MESSIAH's successful verisimilar performance simulation *will never be sufficient* for the occurrence of duplication performance, because verisimilar simulation entails no duplication. Undoubtedly, the performance of Stradivarius' violins is *metaphysically linked* to Stradivari's special manner of manufacturing violins, and the mysterious causal factors mentioned earlier, which prevent the duplication of Stradivarius violins by a mere computer program simulation. What if someone still claimed that *in principle* it

is not impossible to duplicate a Stradivarius violin's performance by its computer simulation? At this point, the distinction between imagination, conceivability and possibility sketched in section 1 helps overcome this objection. From the point of view of imagination, it is indeed possible to conceive of a situation in which scientists duplicate the performance of a Stradivarius violin by simulating it. In the context of metaphysical necessity, the situation is radically different, though. If someone simulated the performance of a Stradivarius violin, the simulation would never perform as an authentic Stradivarius violin, because the origin of the simulation remarkably differs from the *origin* of a bona fide Stradivarius violin's performance. Therefore, the equation of simulation performance and duplication performance in AI is *metaphysically impossible*.

In addition, the Stradivarius violins' case suggests the following consequence: as it is metaphysically impossible to separate a Stradivarius violin's performance from its origin, likewise, human intelligence cannot be isolated

from its origin. And, incidentally, neither can intelligence be encoded into a program. It is metaphysically impossible to isolate linguistic performance from its metaphysical origin, that is, from whatever originates intelligence in the biological dominion. Turing's presupposition that it is possible to separate intelligence from its biological origin is *prima facie epistemically possible — and thus conceivable — but metaphysically impossible*, as intelligence metaphysically depends upon a large number of biological conditions which are *irreproducible* by *pure simulation*. On close analysis, although Artificial Intelligence's simulations involve verisimilitude and deceive interrogators, they will never be sufficient for duplicating intelligence, which explains why verisimilar simulations rule out intelligence duplication in AI and, likewise, why intelligence duplication in AI, was it empirically possible, would involve no simulation whatsoever. Patently, this sort of intelligence duplication would indeed go far beyond the initial project of Artificial Intelligence,

as Turing himself realizes:

> "One might for instance insist that the team of engineers should be all of one sex, but this would not really be satisfactory, for it is probably possible to rear a complete individual from a single sell of the skin (say) of a man. To do so would be a feat of biological technique deserving of the very highest praise, but we would not be inclined to regard it as a case of 'constructing a thinking machine'. This prompts us to abandon the requirement that every kind of technique should be permitted. We are the more ready to do so in view of the fact that the present interest in 'thinking machines' has been aroused by a particular kind of machine, usually called an 'electronic computer' or 'digital computer'." (Turing 1950, p. 42)

In view of this passage, it is more or less obvious why Turing, and Functionalism later, have been accused of endorsing dualism, despite criticizing it. Both Turing and Functionalism claim that it is possible to duplicate the mental abilities of a human being without duplicating its origin, that is, by the pure imitation of linguistic behavior. Claiming so, both theories imply that the mental is separable from the physical.

In the context of refuting Strong Artificial Intelligence and the implausibility that computers understand stories in natural language, Searle objects this point too (Searle 1980 and 1990), claiming that formal symbols, upon which a computer program is based, have no causal powers and, therefore, Artificial Intelligence cannot create intelligence by running programs that solely simulate intelligence. To Searle, as intelligence is caused by the causal powers of the brain, Artificial Intelligence cannot duplicate those powers and, hence, is confined to Weak Artificial Intelligence, or to the design and use of

computer programs as tools to test scientific hypotheses on the human cognitive system. Searle precisely bases his Chinese Room argument upon the impossibility of equating simulation and duplication. However, he does not provide any account to explain *why* one is not justified in claiming intelligence duplication from simulation other than providing instructive examples illustrating why simulation is not duplication, like the following ones:

> "You can simulate the cognitive processes of the human mind as you can simulate rain storms, five fire alarms, diggestion, or anything else that you can describe precisely. But it is just as ridiculous to think that a system that had a simulation of consciousness and other mental processes thereby had the mental processes, as it would be to think that the simulation of digestion on a computer could thereby

actually digest beer and pizza" (Searle 2002, p. 52)

Besides stressing the absent causal component involved in simulation, which is Searle's main point, the Stradivarius violins' example shows why 'machine intelligence' — as it was originally conceived by Turing — and human intelligence cannot clearly be equated in this precise way: both come from two completely different origins and thus cannot be regarded as identical.

Would it be possible to have intelligence duplication in terms different as AI's original project? Searle considers this possibility as follows:

> "[AI's advocates might respond that] Whatever these causal processes are that you say are essential for intentionality (assuming you are right), eventually we will be able to build devices that have these causal processes, and that will be artificial intelligence. ...

> . . . I really have no objection to this reply save to say that it in effect trivializes the project of strong AI by redefining it as whatever artificially produces and explains cognition. . . .
>
> . . . 'Could a machine think?' The answer is, obviously, yes. *We are precisely such machines.* 'Yes, but *could an artefact, a man-made machine, think?*'
>
> . . . If you can exactly duplicate the causes, you could duplicate the effects. And indeed it might be possible to produce consciousness, intentionality, and all the rest . . .
>
> . . . It is, as I said, *an empirical question*." (Searle 1980, pp. 81–82, our italics)

And so is the possibility of duplicating Stradivarius violins' performance by running programs. If possible, the achievement of Stradivarius violins'

performance duplication will eventually be achieved by the scientific reproduction of the very same conditions under which Stradivari, Guarneri, Amati and other artisans crafted their violins in Cremona in the 17th and 18th centuries. This possibility, however, is an empirical question as well.

AI's inability to achieve intelligence duplication by intelligence simulation raises two final questions: What condition must successful simulation meet to be regarded as sufficiently verisimilar and then successfully persuading? What should AI's simulations be like, given that it cannot possibly duplicate intelligence by carrying out pure computerized simulations? Artificial Intelligence ought to simulate intelligence by processes that *have apparently duplicated intelligence*, that is, by imitation processes that *minimally depart from actual intelligence*. Turing certainly misled the tradition, claiming that an indistinguishable performance involves its duplication, at least as far as passing the Turing test is concerned. This view does not indeed follow from intelligence simulation in AI and, *yet*, it is *prima facie conceivable*, at least

in terms of imagining such a possibility. Nevertheless, on closer examination, the duplication of intelligence is not *metaphysically possible* in terms of pure simulation, as the Stradivarius' violins case shows with regard to musical performance.

The situation described is an illustration of a familiar problem in the Philosophy of Mind. In his *Philosophical Investigations* (Wittgenstein 1988 I, 432), Wittgenstein presents us with the following observation and puzzle: "Every sign by itself seems dead. What gives it life? In *use* it is alive. Is life breathed into it there or is the *use* its life?" Wittgenstein's question points directly to one of the problems which is central to the philosophy of language and indeed to philosophy in general, which is the problem of how words and sentences get 'meaning' and 'reference':

> "How can one explain that the signs have the capacity to act as signs, by which mechanisms the physical phenomena underlying the use of language and linguistic acts take on meaning, or how the transposition from the order of the pure external arrangements ... to the order of language takes place." (Ladrière 1984, p. 59).

One possible answer to this question is that words and sentences are the linguistic expressions of (certain) human mental states and as such are characterized by 'intentionality', being that property of mental states by which they can be said to be 'about' something. Giving an account of this 'aboutness', which is characteristic for human language, is considered to be an important problem in any philosophy of language or philosophy of mind. D. Dennett (Dennett 1987) considers, in this respect, three different 'stances', all useful

in predicting the behavior of certain systems. In the *physical stance* one predicts the behavior of a physical system by exploiting information about the physical constituents of the system and the laws of physics. In the *design stance* one predicts the behavior of a system by assuming that it has a certain design, that it is composed of certain elements with certain functions and that it will behave as it is designed to behave under certain circumstances. In so doing, the design stance can safely ignore details of physical implementation of the various imputed functions. Finally, there is the *intentional stance* in which the systems whose behavior one wants to predict are treated as rational agents. They are attributed the beliefs and desires they ought to have given their place in the world and their purpose, and subsequently one predicts that they will act to further their goals in the light of their beliefs.

The research program of 'GOFAI' ('Good Old Fashioned Artificial Intelligence') is not unlike what is being suggested by Dennett's *design stance*.

One of its aims is to make an effort to reduce mental states (to a large extent) to 'functional' or 'computational' ones in order to provide a naturalistic account of these states. As mental states are characterized by intentionality, the theory will have to offer a reductive account of this phenomenon, to the extent that it wants to hold that there are no irreducibly mental properties. Mental states such as believing and desiring are characterized by intentionality and may be said to be 'relational properties' in the sense that they relate people to nonlinguistic entities called propositions (Field 1980, p. 78). Any reductive theory (e.g., a materialist, functionalist or a neurocomputational one) will have to show that these relations are not irreducibly mental but can be reduced to something else, e.g., functional states of the brain and/or the nervous system, which, among other things, implies that no mental state lacking a constitution correctly describable in terms of human neurophysiology could be psychological. But this implies "an absurd ontological imperialism." (Haldane 1988 p. 26). In *Naming and Necessity* Kripke has presented

an argument against any mind-brain identity theory, which is derived from his essentialism and the necessity of identity. It is a consequence of his theory presented there that a substance such as 'gold' should necessarily have the atomic number 79 if that is indeed its *nature*. Also, given that e.g. 'P' is a name of a particular sensation of say, pain, and 'B' is a name of a brain state, an identity theorist would wish to identify 'B' with 'P'. He should then, according to Kripke, hold that ? (P = B). However, on the face of it, we could imagine a possible world where beings with a physical constitution *different from ours* would nevertheless feel pain. In this case, an identity theorist (Helman 1983, p. 156) cannot reply with the suggestion that we are only *imagining* a being with a different physical constitution, who stands in the same epistemic relation we stand in with respect to P but who does not actually feel P, for "to be in the same epistemic situation that would obtain if one had a pain *is* to have a pain. To be in the same epistemic situation that would obtain in the absence of pain is not to have a pain." (Kripke 1980,

p. 152). It therefore seems that this relation between P and B is contingent, in which case P and B are not the same.

However, the move could be made to a weaker physicalist theory, viz. a token-identity theory of mind which implies, among other things, that abstraction is made from a particular physiology. But mental states seem to have 'intrinsic' intentionality and, at first sight, it is not clear how a computational epistemology can ever be said to account for this. The most important problem is that such a theory cannot account for the content (the 'semantics') of those mental states which it is considered to represent. Mental states, such as the belief that Brussels is the capital of Belgium are *about* something, and if in neurocomputational epistemology being in a particular psychological attitudinal state ('to believe') is to be in a particular state or operation on a state this operation is first and foremost a syntactical operation. The functional states are syntactically defined processes, and they require a physical

substratum, as do the psychological states which they are supposed to represent somehow. In the case of a psychological state such as belief, at least two things should be required from such a theory if one wants to account for the intentionality of the mental state involved. First, we would need a psychological theory, H, containing primitive predicates such as B (believe) where one assumes that belief is a relation between an organism (which is in a specific state) and a proposition expressing the content of the belief. Belief would then be a functional relation associated with some theory in which the term 'believe' occurs. Second, we can now say that B is a functional relation F of belief from GOFAI's computational epistemology F, T(F), and that an individual x is in the functional relation of belief to p (a proposition) at a certain moment m, T(m), if there is a physical property S (the brain state) which x bears to p at a particular moment T(m). In other words, if there is no physical relation (a state of affairs in the neural network) then x cannot stand in the required relation B. This functional relation is not, in itself, a

physical relation; it would be represented within the theory by a property of functional states, but it would relate an organism to a proposition. But, then again, one might insist on there being some physical relation which realizes just this, and which relates the organism to the proposition. So what one is left with is that one has to show that there are physical relations (which are non-functional) between people and propositions, and *which would account for the intentionality of the proposition*, and that is not what is done in any functionalist epistemology as it now stands. Here also, therefore, one may say that the duplication of intentionality is not metaphysically possible in terms of pure simulation. And, nor is it required, according to the notional task of Artificial Intelligence.

Conclusion: The verisimilitude of Artificial Intelligence

As we have seen, the comparison between fiction and Artificial Intelligence fleshed out their commitment to persuading readers and interrogators by verisimilar imitations. Nevertheless, whereas the former verisimilarly simulates the actual world in the process of reading fiction, the latter verisimilarly simulates intelligence by the indistinguishable performance of man and machine. Such a comparison additionally indicated that neither discipline requires duplication; hence, Artificial Intelligence need not duplicate actual human intelligence to succeed and accomplish its notion task. To arrive at this conclusion, we have explained how both the make-believe processes of fiction and Artificial Intelligence involve mimicking processes. Nevertheless, the simulation process necessarily requires verisimilitude, a concept that was characterized in terms of Ryan's principle of minimal departure. Based upon Kripke's metaphysical notion of origin and its link to identity,

the Stradivarius violins' case illustrated why the equation of intelligence simulation and its replication is *metaphysically impossible* in terms of pure simulation. Similarly, as the verisimilar intelligent performance of machines does not entail human intelligence duplication, and simulation excludes duplication, Artificial Intelligence can only yield verisimilar simulations, when minimally departing from actual intelligence. Finally, we also analyzed another way in which Artificial Intelligence is able to carry out simulation, but unable to metaphysically duplicate mental states: given the functional character of any GOFAI theory about mental life, which attempts to reduce intentionality in functional terms, mental states are to be explained in terms of relational properties between organisms and propositions. In so doing, any functionalist theory explains the Intentionality of *belief* by providing a theory which reduces *functionally* the relation between the organism, which is in a given brain state S at a moment m when holding the belief, and the content of this belief. Nevertheless, since these functional relations expressed

by the theory, which need a material realisation, are *not* physical relations, the functional explanation cannot elucidate intentionality properly. That is, this reduction cannot account for the relationship between the organism and the proposition and its content. This, again, shows why simulation cannot be duplication. Consequently, duplication by simulation remains conceivable and would still be impossible. Although the creation of intelligence by the duplication of its origin is a plausibly achievable empirical task, advocating this possibility in the case of AI only trivializes its original project. This is, indeed, unnecessary, given the notional task of Artificial Intelligence.

Center for Logic, Philosophy of Science and Language

Institute of Philosophy

Catholic University of Leuven

Kardinaal Mercierplein 2

3000 Leuven

Belgium

E-mail: Roger.Vergauwen@hiw.kuleuven.ac.be

REFERENCES

Block, N. (1990): "The Mind as the Software of the Brain." In: D.N. Osherson & H. Lasnik (eds.) *Thinking: An Invitation to Cognitive Science*, Vol. 3. Cambridge (Mass.), MIT-Press, pp. 377–425.

Carnap, R. (1946): "Modalities and Quantification", *Journal of Symbolic Logic* 11, 33–64.

Carnap, R. (1947): *Meaning and Necessity: A Study in Semantics and Modal Logic*, Chicago, University of Chicago Press.

Cleland, C. (1993): "Is the Church-Turing thesis true?", *Minds and Machines* 3, 283–312.

Chalmers, D. (2002): "Does Conceivability Entail Possibility?" In: T. Szabo Gendler & J. Hawthorne (eds.) *Conceivability and Possibility*. Oxford, Oxford University Press, pp. 145–200.

Copeland, B.J. (2000): "The Turing Test." In: James H. Moor (ed.) *The Turing Test: The Elusive Standard of Artificial Intelligence*. Dordrecht, Kluwer Academic Publishers, pp. 1–21.

Copeland, B.J. (2002): "The Genesis of Possible World Semantics", *Journal of Symbolic Logic* 31, 99–137.

Dennett, D. (1987): *The Intentional Stance*. Cambridge (Mass.), MIT-Press.

Field, H. (1980): "Mental Representation." In: N. Block (ed.) *Readings in the Philosophy of Psychology*. Cambridge, (Mass.) Harvard University Press, pp. 78–114.

French, R. (2000): "The Turing Test: The First Fifty Years", *Trends in Cognitive Science* 4, 115–122.

Gough, Colin (2000): "Science and the Stradivarius." In: *Physics World*, at: http://physicsweb.org/articles/world/13/4/8/1.

Haldane, J. (1988): "Psychoanalysis, Cognitive Psychology, and Self-Consciousness." In: P. Clark et al. (eds.) *Mind, Psychoanalysis and Science*. Oxford, Blackwell, pp. 113–139.

Helman, D. (1983): *A Perspective on the Philosophy of Saul Kripke*. Ph.D Harvard (University Microfilms).

Hodges, A. (1992): *Alan Turing: The Enigma*. London, Vintage.

Hume, D. (1978): *A Treatise of Human Nature*. Edition of L.A. Selby-Bigge & P.H. Nidditch. New York, Oxford University Press.

Kripke, S. (1980): *Naming and Necessity*. Oxford, Blackwell.

Ladrière, J. (1984): "Signification et Significance", *Synthese* 59, 59–67.

Lewis, D. (1979): "Possible Worlds." In: Michael J. Loux (ed.) *The Possible and the Actual: Readings in the Metaphysics of Modality*. Ithaca, Cornell University Press, pp. 182–89.

Mele, A. (1996): "Agency and Mental Action", *Philosophical Perspectives* 11, 231–249.

Putnam, H. (1975): "The Meaning of 'Meaning'." In: K. Gunderson (ed.) *Language, Mind and Knowledge*. Minnesota Studies in the Philosophy of Science, VII. Minneapolis (Minn.), University of Minnesota. Reprinted in: Hilary Putnam, *Mind, Language and Reality. Philosophical Papers, Volume 2*. Cambridge, Cambridge University Press, pp. 215–271.

Putnam, H. (1992): "The Project of Artificial Intelligence." In: Hilary Putnam, *Renewing Philosophy*. Cambridge (Mass.), Harvard University Press, pp. 1–18.

Rescher, N. (1979) "The Ontology of the Possible." In: Michael J. Loux (ed.) *The Possible and the Actual: Readings in the Metaphysics of Modality*. Ithaca, Cornell University Press, pp. 166–181.

Ryan, Marie-Laure (1991): *Possible Worlds, Artificial Intelligence and Narrative Theory*. Bloomington (Ind.), Indiana University Press.

Saygin, A.P. et al. (2000): "Turing Test: 50 years later." In: James H. Moor (ed.) *The Turing Test: The Elusive Standard of Artificial Intelligence*. Dordrecht, Kluwer Academic Publishers, pp. 23–78.

Schlick, M. (1932/1933): "Positivism and Realism.", *Erkenntnis III*, 1932/33. Reprinted in: Richard Boyd et al. (eds.) *The Philosophy of Science*. Cambridge (Mass.), MIT-Press, pp. 37–55.

Searle, J. (1980): "Minds, Brains and Programs", *The Behavioral and Brain Sciences* 3, 417–424. Reprinted in: Margaret Boden (ed.) *The Philosophy of Artificial Intelligence*. Oxford, Oxford University Press, pp. 67–88.

Searle, J. (1990): "Is the Brain's Mind a Computer Program?", *Scientific American*, January 1990, 20–25.

Searle, J. (2002): "Twenty-One Years in the Chinese Room." In: J. Preston & M. Bishop (eds.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford, Oxford University Press.

Thomasson, R. (2003): "Logic and Artificial Intelligence." In: *Stanford Encyclopedia of Philosophy*, at:

`http://www.science.uva.nl/ seop/entries/logic-ai/.`

Turing, A.M. (1948): "Intelligent Machinery", National Physical Laboratory Report. In B. Meltzer & D. Michie (eds.) *Machine Intelligence* 5, Edinburgh: Edinburgh University Press, pp. 3–23.

Turing, A.M. (1950): "Computing Intelligence and Machinery", *Mind* LIX, no. 2236, (Oct. 1950), 433–460. Reprinted in: Margaret A. Boden (ed.) *The Philosophy of Artificial Intelligence*. Oxford, Oxford University Press, pp. 40–66.

Turing, A.M. (1951): "Can Digital Computers Think?" In: K. Furukawa et al. (eds.) *Machine Intelligence* 15. Oxford: Oxford University Press, pp. 42–60.

Wittgenstein, L. (1988): *Philosophical Investigations*. Oxford, Blackwell.

Yablo, S. (2002): "Coulda, Woulda, Shoulda." In: T. Szabo Gendler & J. Hawthorne (eds.) *Conceivability and Possibility*. Oxford, Oxford University Press, pp. 441–92.