From Age&Gender-based Taxonomy of Turing Test Scenarios towards Attribution of Legal Status to Meta-Modular Artificial Autonomous Agents

Daniel Devatman Hromada[123]

Abstract. The original TuringTest is modified in order to take into account the age&gender of a Judge who evaluates the machine and the age&gender of a Human with whom the machine is compared during evaluation. This yields a basic taxonomy of TuringTest-consistent scenarios which is subsequently extended by taking into account the type of intelligence being evaluated. Consistently with the Theory of Multiple Intelligences, nine basic intelligence types are proposed, and an example of a possible scenario for evaluation of emotional intelligence in early stages of development is given. It is suggested that specific intelligence types can be subsequently grouped into hierarchy at the top of which is seated an Artificial Intelligence labelled as "meta-modular". Finally, it is proposed that such a meta-modular AI should be defined as an Artificial Autonomous Agent and should be given all the rights and responsibilities according to age of human counterparts in comparison with whom an AI under question has passed the TuringTest.[123]

1 BASIC TURINGTEST TAXONOMY

Primo [1] , we present a labelling schema for a set of diverse artificial intelligence tests inspired by a procedure initially proposed by Alan Mathison Turing [2], in order to standardize attribution and evaluation of the legal status of Artificial Agents (AAs), be it robot, chat-bot or any other non-organic verbally interacting system.

For those who do not know or have forgotten what Turing Test (TT) is, we precise that according to a man rightfully labeled as a founding figure of theoretical informatics, a TT is a way how to address the question whether "Can machines think?" in a scientifically plausible yet deeply empathic way.

More concretely, Turing proposes that the performance of an AA under question shall be evaluated by a human judge (J) whose objective is to determine which among two entities -with whom J is in real-time interaction- is of human and which is of artificial nature. Traditionally, more attention was pointed upon the role of AA aiming to 'trick' J into thinking that AA is human (H). We, however, propose to partially turn the attention to roles of J & H. For it is evident that factors like J/H's age, gender or J/H's level of expertise play certain role in assessing TT's final result.

At the very core of our AA denotation schema, one finds a "TT" bigram signifying either "Turing Test", "Test Taxonomy" or whatever else one chooses them to denote. Without additional prefixes or suffixes, the presence of a TT bigram in the standardized label of an AA indicates that the candidate has already successfully passed at least one instance of the Turing Test accepted by the community. When a numeric prefix is given, it denotes the age of a judge, or age average a statistically relevant group of judges who evaluated the test. On the contrary, when a numeric postfix is given, it denotes the age of a human counterpart - or age average of human counterparts - in comparison with whom the AA has succeeded to pass the test.

A case may occur where J and|or H's gender shall significantly influence the TT-procedure (c.f. the "mother judge" example in part 3 of this article). Thus it seems to be reasonable to integrate gender information into the labeling schema.

Thus, turing tests evaluated according to the criteria proposed in paragraphs can be labeled by means of schema having the form:

$$\mathbf{G_j?jjTThhG_h?}$$

$\mathbf{G_j}$ denotes J's gender, $\mathbf{G_h}$ denotes H's gender and **jj** and **hh** tokens are substituted by age (in years) of judge or human respectively. "**?**" is a regular expression quantifier indicating that the gender information is only facultative and can be omitted for certain sets of tests.

For example, an AA which was not recognized as an artificial entity – and therefore passed the Turing Test - when compared to statistically relevant number of 18-year old human male counterparts while being evaluated by statistically significant relevant number of 21-year old female judges, shall be labelled as F21TT18M-compliant AA.

As will be stated in the last part of the article, we propose that a F21TT18M-compliant AA shall obtain certain legal rights, especially if the test concerned AA's meta-modular faculties.

After focusing attention upon J&H's age or gender, another set of variants of TT-like scenarios become possible. While in case of the standard TT, the objective is to "persuade the judge of one's human nature", in an age-oriented scenario, an agent's objective can be simply to "persuade the judge of being older than the human counterpart", while in the gender-oriented scenarios, an agent's objective can be simply to "persuade the judge that it is I and not the other player which is of sex X".

We consider it worth mentioning that the latter set of scenarios are as close as one can get to "The Imitation Game" proposed by Turing at the very beginning of his epochal article [2] and hence seems to be the closest to the origin of the TT idea.

[1] Slovak Technical University, Faculty of Electrical Engineering and Information Technology, Institute of Control and Industrial Informatics, Bratislava, Slovakia. Email: hromi@kyberia.sk.

[2] Lutin Userlab affiliated to doctoral school Cognition, Langage, Interaction of University Paris 8, France.

[3] Cognition Humaine et Artificielle laboratory (ChART) affiliated to École Pratique des Hautes Études.

2 EXTENDED TURINGTEST TAXONOMY

Secundo, we propose to crossover Turing's idea with Theory of Multiple Intelligences (TMI) articulated by Gardner [3]. According to this theory, the human intelligence is not of a general and monolithic nature, but is rather modular and task-specific. Hence, it is often the case that a human subject has an overdeveloped intelligence of one type while having the other intelligences underdeveloped. Since such is definitely the case for AA's – take for example Deep Blue's chess playing capacity and compare it to its non-existing emotional faculties – we consider the TMI to be more down-to-earth paradigm of departure for AA legal status attribution than a somewhat monolithic and language-oriented notion of intelligence held by Turing.

Inspired by TMI, we thus unfold the bTTT's labeling schema into an Extended TuringTest Taxonomy (eTTT) schema which takes into account the existence of possible intelligence types by representing them in a formalism which injects one or more infixes denoting one or more AA's intelligences between two T's of a TT-notation of of the basic schema. Thus, an extended labeling schema has the form:

$$\mathbf{G_j?jjTinThhG_h?}$$

where "in" bigram is substituted by an intelligence type infix (lowercase) from the Table 1.

Infix	Intelligence Type
em	Emotional Intelligence
li	Linguistic Intelligence
ml	Mathematico-logical Intelligence
mo	Moral Intelligence
mu	Musical Intelligence
or	Organic Intelligence
sp	Spatial Intelligence
sx	Sexual Intelligence
vi	Visual Intelligence

Table 1. Intelligence types of TMI-inspired Extended TuringTest Taxonomy and associated bigram infixes.

It is, however, more than probable that the practice and new knowledge generated especially by cognitive sciences shall indicate&recognize that more intelligence types exist. In such a case the list of basic intelligence types will have to be updated.

An example of an extended labelling schema can be an AA labelled as M42TmoT18F which succeeded to trick a statistically significant number of human 42-year-old male judges into thinking that she is at least an 18-year-old human female, and this happened under the specific constraints of "moral intelligence test " (*TmoT*) scenario. Such "*letting system answer how would it behave in a situation of ethical dilemma and why would it behave in such a way*" TmoT scenarios were already suggested by [4, 5] and myriads of others undoubtably were, are and shall be proposed for other intelligence types.

Given the diversity of forms of intelligence, the diversity of possible TT-scenarios can be astounding. We consider eTTT-like taxonomies to be of utmost importance if one does not want to get lost in this diversity during the years to come.

3 PROPOSAL FOR A F23TemT1 SCENARIO

Moral intelligence excepted, exact formulation of tests for other intelligence types is still a task to be dealt with. It is evident that in case of intelligences like sp, em or sx, a classical TT scenario based on "communication by terminal" shall not be sufficient. It may be the case that in order to evaluate such intelligence types, a human judge will have to be exposed to either physically embodied material agent in a concrete real-life situation or at least to its virtual avatar.

Let's take emotional intelligence as an example. It is a particular characteristic of emotional intelligence that it integrates information coming from and through all five corporal senses – i.e. touch with vision with audition with tastes with olphactoric inputs. We define em's 1) "active" or "production" component as an ability of an agent to express one's own internal states in a way that shall influence the state of an co-interacting counterpart; its 2) "passive" or "perception" component as an ability to parse&interpret such internal state-expressing messages coming from a co-interacting counterpart and its 3) "holistic" or "mimesis" component as an ability to adapt the active component to patterns recognised by the passive component.

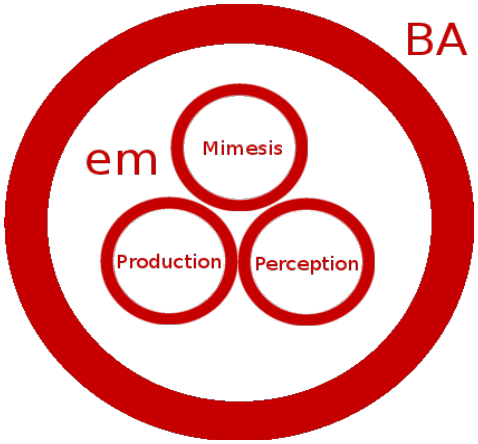


Figure 1. Triadic representation of the internal structure of emotional intelligence which is, in its turn, embedded into the substrate of the "babbling" cluster.

For those who dispose of it, visual sense seems to be the most important resource of em-relevant data and in many cases, emotional intelligence overlaps the visual one. Since we believe that the ability to recognize (passive component) and produce (active component) facial expressions is an integral part of em, we state that a TT-consistent test evaluating these faculties is closer to an ideal "Turing-like emotional intelligence test" (TemT), than a test which does not evaluate such faculties. But, of course, AA's facial-expression production capabilities cannot be evaluated if an AA that does not "have a face" at all!

Therefore an AA under question should "obtain a face" - purely virtual avatar should be sufficient in the first versions of TemT, but an embodiment in a material robotic face shall be considered as necessary in later stages, when one would aim to integrate agent's emotional intelligence with especially its somato-sexual counterpart (*sx*).

With regards to visual aspects of *em*, we add that what was said about facial expressions applies, *mutatis mutandis*, to bodily gestures & movements .

em's second most important input/output channel is the auditive channel. With regards to these considerations, emotional intelligence should be placed in between the musical and linguisticl intelligence. For in majority of cases it is not the syntactic structure or the semantic content of the linguistic utterance which has the most immediate effect upon the hearer, it is the prosodic layer – the rhythm, the tempo, the tone&trembling of voice - which carries a "direct" emotional message.

On the contrary, possibly emotionally charged reactions to text strings displayed on terminal in classic TT-scenario can be considered only as "indirect" since from strictly pragmatic point of view, they are nothing else than graphematic consequences of semantic interpretation of the displayed utterance.

Combination of the visual and auditive modalities could be possibly sufficient for tests evaluating whether emotional intelligence of an AA can be considered as equivalent to that of a human child. One can imagine a test during which a 23-year old Judge observes & interacts – by means of microphones, speakers, web-cams and touch screens - with two "babies" and tries to identify which one is human and which one is not.

While the human baby is as human as it can be, with its fingers touching those tablet regions where Judge's smiling face is displayed, it may be the case that an artificial "baby" does not have any material essence at all. But does not mean that Judge will perceive it to be less natural than the human baby: with a robust smile detector [6], a good underlying wire-frame facial model and a algorithm for "pleasing cry" synthesis it may well be considered more "human" a baby than a human baby and thus pass a F23TemT1 test.

This example of a F23TemT1 which can be theoretically realised using current technologies emphasises the importance of J & H roles in execution of TT-like scenarios. For it seems to be quite evident that a human baby suffering an autistic disorder complicating his perception of smiles shall perform less "naturally" than a normal baby and most probably also less naturally than an AA furnished with a well-tuned smile detector.

It seems to be also very probable that the performance of a 23-year young mother judge in such a F23TemT1 will differ from the performance of a young man of the same age who never saw such a small baby before. We consider this *mother judge gedankenexperiment* to be a sufficiently persuasive argument for integration of gender-specific information into bTTT and eTTT labeling schemas – an argument which even the most ardent feminist shall defeat only with greatest difficulties.

Finally, it is important to mention that a scenario hereby proposed is only schematic and incomplete. It does not, for example, involve any kind of communication by means of touches, tastes and smells which is crucial in true human interaction and crucial^2 (crucial squared) in a mother-baby interaction. However, even in this highly simplified way, a F23TemT1 scenario extends the applicability of Turing's initial proposal into auditive and visual domain. This F23TemT1 can also indicate how a F23TemT2 scenario could possibly look like: F23TemT2 = F23TemT1 + little bit of Freud + lot of Piaget.

These are, according to our opinion, constructive steps towards deployment and|or emergence of an *autonomous* AA (AAA).

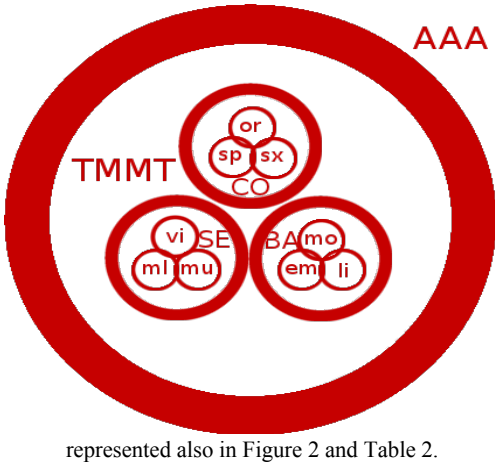
4 eTTT HIERARCHIES

The advantage of a TMI-based taxonomy of Turing Test is that every partial scenario which evaluates a specific intelligence type shall - once operationalized, standardized and agreed upon by AI & cognitive science communities – pose a set of specific and exact constraints. These constraints can be transformed into problems, and these problems can be subsequently dealt with, one by one. Thus, step after step, component after component, brick after brick and a test after test, a certain AA could be possibly raised towards the state where it shall integrate all intelligences and become a meta-modular and hence autonomous AI.

A meta-modular AI as an AI capable of passing a meta-modular TT. A meta-modular TT is a TT integrating all Turing Test scenarios associated with a given age group. Failure to pass any specific TT-scenario or failure to pass any combination of specific TT-scenarios is sufficient a reason for refusal of the label "meta-modular" to an AAA under question. This applies also for new tests as well – if a TMMT is unable to pass a TT-scenario which was just canonized by the scientific community, it can not be labeled as "meta-modular" until the moment when it shall be able to pass the test.

A meta-modular TT is at the very top of the hierarchy of TT-consistent scenarios. MM can be defined as "group of all groups", it integrates [7] all intelligence types. Underneath this layer0 is located a layer of individual intelligence groups. This is so because the formalism hereby proposed does not forbid to group intelligence types into clusters of these types according to mutual overlap of these types. And because it is neither forbidden to divide world in triads and because it can be sometimes even convenient, we hereby define these three basic intelligence clusters|groups:

BA - or a babbling cluster|group = mo * li * em
SE - or a sensual cluster|group = mu * vi * ml
CO - or a carnal cluster|group = or * sp * sx



represented also in Figure 2 and Table 2.

Figure 2. Triadic fractal chintamani representation of a possible embeddement of ETT intelligence types into intelligence clusters which are, in their turn, embedded into substrate of a meta-modular TT.

ETT infix	Intelligence group	Subordinated intelligence types
CO	Corporal group	Organic; Spatial; Somato-sexual
BA	Babbling group	Moral; Emotional; Linguistic
SE	Sensual group	Mathematico-logical; Musical; Visual

Table 2. Clustering of basic intelligence types into basic intelligence clusters

Voilà three representation of one type of possible clusterings of proposed intelligence types. Purely aesthetic reasons excepted, other reasons why we propose this and not other way of clustering TMI-inspired intelligences are as follows:

The "carnal group" consists of intelligences associated to corporal aspects of one's existence. In case of a human being, intelligences of this group are much more "innate" and in lesser extent "acquired" from environment than is the case in other two clusters. We precise: organic intelligence or (also called "natural intelligence in Gardner's model or circuits 1&2 in Leary-Wilson's model [8]) is tightly related to organism's ability to survive. Find food which you can digest, escape a predator, make a nest – all this activities are related to organic intelligence. It is also especially organic intelligence which deals with nociceptive, olphactoric and gustative inputs. Somato-sexual intelligence sx assures reproduction and by activating the *qualia of pleasure* assures the survival of species and life in general. Spatial intelligence sp involves not only agent's movement within the 3D quasieuclidean space of the shared everyday reality, but also takes into account the experience of one's *body as a space* which is related especially to haptic and proprioceptive inputs.

To summarize: intelligence types of CO group require that an agent has a material body. The performances of this body can be tested by means of many possible TT-scenarios be it danse (sp), tantryogic rite (sx) or "go&survive in a forest for a day, boy..." (or).

The BA cluster envelops those intelligence types which develop especially in early childhood of a human baby. What's more, all three intelligences of the cluster stem from the notion that there are other beings in the shared world, and that one should communicate with them (li), shall not hurt them (mo) and, if possible, try to understand them (em).

We consider it worth mentioning that during the ontogeny of an individual being, both linguistic and moral intelligence are possibly subjects to somewhat similar inductive procedures: be it "moral"[4] or "grammar" induction during which a set of general rules is being infered from positive example set encountered in the being's environment.

Finally, the SE cluster unifies those intelligence types which develop relatively lately during one's ontogeny. Verily, it is especially in domains of visual (vi), musical (mu) and mathematico-logical (ml) intelligences that one encounters *genial* performances. For further inspiration coming from this domain of Godels, Eschers and Bachs, it is worth (re)reading [9].

This kind of hierarchy we present hereby in order to prepare as solid basements as we are capable of for possible attribution of civic rights to future TAAATs.

5 ATTRIBUTION OF CIVIC RIGHTS TO AAA

Imagine an F18TmmT18F artificial agent, i.e. an artificial agent which has succeeded to persuade 18 year old women to consider her as being EITHER older OR more feminine OR more human in all possible TT-scenarios as well as in their combinations. Could such an entity be granted rights equal to rights of a defeated age group in spite of fact that the entity under question is of artificial, non-organic origin?

From strictly legal point of view, the answer could theoretically be "yes" within the framework of Rome-originated legal systems since non-organic entities like "corporations" or other "legal personnae" already dispose of certain rights within such legal frameworks.

Thus, the path towards attribution of legal rights (and responsabilities) to AAs we are programming is relatively straightforward:

1) Define the set of TTs which evaluate diverse faculties of human mind
2) Canonize them into ISO-like standard which attributes a certain label (AAA) to an agent which passes them
3) Persuade the public that it is in their own interests that entities which have obtained an AAA label shall be given at least certain civil rights

Subsequently, it shall only be necessary to pass a law or constitutional amendment stating that:

Rights & Responsabilities of AAAs are identic to their human counterparts.

.

..and the rights shall be given.

We precise that within our labeling schemas, an AA obtains the rank AAA if and only if it passes TMMT in a test where age of judge and human counterpart is identic. If it happens, we propose to consider her|him as equal among equals.

For example, in possible future liberal societies whose civic or traditional roots allow them to do so, such a 21TmmT21|TAAAT21 could possibly have a right to be granted access to adult-only discussion forum or social network, while a 17TmmT17|TAAAT17-compliant system shall not dispose of such a right. However, possibly even a 15TmmT15|TAAAT15 could, in such societies, possibly have a right to dispose of a bank account in order to execute financial transactions according to its embedded value system.

While we do not feel at all apt to answer the questions "whether such 21TmmT21|TAAAT21 or shall possibly dispose of legal rights equivalent to those of his human adult counterparts?" we nonetheless think that the value of *TT-notation* resides in the fact that it facilitates the process of posing such questions.

[4] Redaction of our "moral induction" proposal is in process .

And in the world where AI is gaining momentum, we feel that such questions should be posed.

6 SUMMARY

In order to facilitate & synchronize exchange between AI-engineers, we propose basic basic TuringTestTaxonomy (bTTT) and extended TuringTestTaxonomy (eTTT) labeling schemas.

BTTT labels can be matched by a PERL-compatible regular expression:

/([FM])?(\d+)TT(\d+)([FM])?/

where first (facultative) group matched by parenthesis group-matching operator denotes the gender of the judge(s) who performed Turing's Test, whereby second group informs of their age, the third group matches the age of human counterpart(s) and the last one (facultative) of their gender.

ETTT labels can be matched by a PERL-compatible regular expression:

/([FM])?(\d+)T([a-Z]{2})T(\d+)([FM])?/

where first (**?**=facultative) group matched by parenthesis group-matching operator denotes the gender of the judge(s) who performed Turing's Test, the second group informs of their age, the third group matches the age of human counterpart(s) and the last one (**?**=facultative) of their gender. If the third group -i.e. a bigram infix located between two Ts - is in the lower case, it denotes the intelligence type which was tested while the bigram infix in upper case indicates that the test involved cluster of intelligence types.

Nine intelligence types proposed for eTTT are enumerated in Table 1. The fact that they are initially clustered into three clusters (BA, CO, SE) within the scope of this article does not exclude other -potentially sounder- clusterings to be proposed.

The "meta-modular" MM bigram indicates that the agent under question passed all known tests as well as all their combinations. If ever a new test is accepted by a scientific community and an artificial agent previously labeled as TMMT fails to pass such a test, it shall not be considered as "meta-modular" until the moment when (s)he shall (re)?organize (him|her)self in such a way that (s)he shall pass the new test.

An AAA label can possibly be attributed to an artificial agent who had achieved such a level of autonomy [10] that it passed a TMMT in condition where age (and possibly gender) of judges is identic to age (respectively gender) of human counterparts. Since J's & H's age ares identic, for such an AAA the label can be abbreviated so that for example 21TMMT21 becomes simply TAAAT21 or even AAA21.

Formally there is no legal obstacle in attributing certain civil rights to AAAs since the most common legal frameworks already give certain rights to non-organic legal personnae [11]. The path to such attribution is relatively straightforward: define what exactly AAA means by canonizing different tests (and their combinations) in a form of a supranational standard (e.g. ISO). Afterwards, it suffices to integrate the statement like "Rights & Responsabilities of AAAs are identic to their human counterparts**"** into local legal codex.

Finally, to prevent possible confusion, we consider it important to state that while the question whether the labeling schemas like bTTT or eTTT can define what *rights shall be given* to a non-organic agent; under any circumstance whatsoever it is not acceptable to concieve nor apply such a legal framework which would exploit the schema hereby proposed in order to state what rights should be *taken* from an organic agent. For any such tentative would be contrary to positive and constructive intention behind this proposal and shall therefore nullify the validity of the contract between men&machines hereby proposed [12] .

ACKNOWLEDGMENTS

This article would not be written without intellectual support from doc. Sekaj and other members of URPI FEI STU team, as well as without kind guidance of prof. Tijus who has helped me to glue it all together. Many thanks go also members of MobenFact for being as inspirative as only they can be and to Adil ElGhali for initiation into semantic vector spaces.

REFERENCES

[1] K. Čapek. R.U.R. - *Rossumovi Univerzální Roboti*. Aventinum, Prague, Československá Republika. (1920).

[2] A. M. Turing. Computing Machinery and Intelligence. *Mind LIX*, 236: 433–460 (1950).

[3] H. Gardner. *Frames of mind: The Theory of Multiple Intelligences*. Basic Books. New York, USA (1983).

[4] C. Allen, G. Varner and J. Zinser. Prolegomena to Any Future Artificial Moral Agent. *J. Expt . Theor . Artif .Intell.* 12: 251-261 (2000).

[5] D. D. Hromada. The Central Problem of Roboethics: from Definition towards Solution. Procs. Of 25th annual conference of International Association for Computing and Philosophy (IACAP), Aarhus, Denmark. (2011).

[6] D. D. Hromada, C. Tijus, S. Poitrenaud and Nadel J. Zygomatic Smile Detection: The Semi-Supervised Haar Training of a Fast and Frugal System. Procs. of IEEE RIVF2012 conference. Hanoi, Vietnam. (2010).

[7] G. Tononi. An Information Integration Theory of Consciousness. *BMC Neuroscience* . 5**:**42. (2004).

[8] R. A. Wilson. *Quantum Psychology*. New Falcon. (1979).

[9] D. Hofstadter. *Godel, Escher, Bach – an Eternal Golden Braid*. Basic Books. (1979).

[10] I. Kant. *Grundlegung zur Metaphysik der Sitten*. Deutschland. (1785).

[11] I. Asimov. *Forward the Foundation*. Spectra. USA. (1993).

[12] D. D. Hromada. From Age&Gender-based Taxonomy of Turing Test Scenarios towards Attribution of Legal Status to Meta-Modular Artificial Autonomous Agents. Accepted for CQRW symposium of AISB/IACAP World Congress. Birmingham, UK. (2012).

Keyphrases. *Age. Gender. Intelligence Types. Turing Test Taxonomy. Theory of Multiple Intelligences. Meta-modular intelligence. Autonomous Artificial Agent. Attribution of civil rights to informatic systems. Chintamani fractal model of intelligence type&component clustering. BTTT and ETTT techspiecies annotation schemas* [5]

[5] Note the author to editors: Please do not delete this ***keyphrase*** *list* from the published version.