e.publica
public law journal

# Touching the core or scratching the surface? Background theoretical aspects of similarity and legal analogy in Artificial Intelligence

# O ponto fundamental ou uma questão lateral? Aspectos Teóricos sobre Semelhança e Analogia Jurídica na Inteligência Artificial

**Pedro Moniz Lopes**

ISSN 2183-184x

Vol. 11 No. 1
maio 2024
e-publica.pt

Com o apoio de:

Fundação para a Ciência e a Tecnologia

*e-Publica* **Vol. 11** No. 1, maio 2024 (004-033)

# TOUCHING THE CORE OR SCRATCHING THE SURFACE? BACKGROUND THEORETICAL ASPECTS OF SIMILARITY AND LEGAL ANALOGY IN ARTIFICIAL INTELLIGENCE [i]

O PONTO FUNDAMENTAL OU UMA QUESTÃO LATERAL? ASPECTOS TEÓRICOS SOBRE SEMELHANÇA E ANALOGIA JURÍDICA INTELIGÊNCIA ARTIFICIAL

**PEDRO MONIZ LOPES**[1, 2]

https://orcid.org/0000-0003-4797-354X
Lisbon Legal Theory | Lisbon Public Law | ICJP - FDUL
plopes@fd.ulisboa.pt

**Abstract:** This paper addresses theoretical discussions surrounding similarity and legal analogy, both in general and as particularly applied to artificial intelligence. It begins by outlining the main reasons why it is said that current AI systems are far from deploying the capability of forming humanlike abstractions or analogies. Problems of analogical reasoning then are divided into general problems of competing background theories specific problems of applicability of those theories in the context of AI (or the feasibility thereof). It recovers some fundamental definitions for the debate of analogy and proceeds to address the inferential model by singling out the specific problems of normative analogies. The paper intends to lay down the fundamental normative and psychological discussions as a general problem, in order to apply them to the specific context of AI. Tversky's directionality and diagnosticity principles are framed as some good insights for the current debates about "sticky" stereotypes and discriminatory treatments attributed to AI systems. Lastly, the paper focuses on the concepts of "relevant similarity" and "sufficient similarity" in AI. It stresses that perceptual processes still prove important to current research on AI and it suggests that programming Rosch's two psychological principles of categorization into AI Systems would helpfully provide for flexible criteria to finetune the level of abstraction of categories formed in analogy-making.

**Keywords:** Similarity, analogy, artificial intelligence, categories, perceptions

**Resumo:** Este artigo aborda os debates teóricos em torno da semelhança e da analogia jurídica, tanto em geral como na inteligência artificial (IA). Começa por delinear as principais razões pelas quais se diz que os atuais sistemas de IA estão longe de desenvolver a capacidade de formar abstrações ou analogias semelhantes às humanas. Posteriormente, os problemas de raciocínio analógico são divididos em problemas gerais de teorias concorrentes e problemas específicos de aplicabilidade dessas teorias no contexto da IA. O artigo recupera algumas definições

1. Assistant Professor at Lisbon University – Law School. Member of LLT – *Lisbon Legal Theory*. Effective Integrated Researcher of Lisbon Public Law – Coordinator of the Research Group on Theory and Philosophy of Law. https://pedromonizlopes.academia.edu/

2. I am indebted for the insightful comments from Alessio Sardo and from the participants of the VI Lisbon Meeting on Legal Theory (on the topic of "Artificial Intelligence & Judicial Decision") that took place at the Lisbon Law School on 30 June 2023. The usual disclaimer applies: mistakes or shortcomings are my own.

fundamentais para o debate sobre a analogia e aborda o modelo inferencial, destacando as dificuldades específicas das analogias normativas. Pretende-se definir as discussões normativas e psicológicas fundamentais como um problema geral, a fim de as aplicar ao contexto específico da IA. Os princípios da direcionalidade e de diagnosticidade de Tversky são enquadrados como ideias úteis para os debates atuais sobre estereótipos recalcitrantes e tratamentos discriminatórios atribuídos aos sistemas de IA. Por último, o artigo centra-se nos conceitos de "semelhança relevante" e "semelhança suficiente" na IA. Sublinha-se que os processos perceptivos continuam a ser relevantes para a investigação atual sobre a IA e sugere-se que a programação dos dois princípios psicológicos de categorização de Rosch nos sistemas de IA poderia fornecer critérios flexíveis para afinar o nível de abstração das categorias formadas na criação de analogias.

**Palavras-chave:** Semelhança, analogia, inteligência artificial, categorias, percepções.

> "There is no property ABSOLUTELY essential to one thing. The same property which figures as the essence of a thing on one occasion becomes a very inessential feature upon another. Now that I am writing, it is essential that I conceive my paper as a surface for inscription.... But if I wished to light a fire. and no other materials were by, the essential way of conceiving the paper would be as a combustible material... The essence of a thing is that one of its properties which is *so important for my interests* that in comparison with it I may neglect the rest... The properties which are important vary from man to man and from hour to hour... many objects of daily use – as paper, ink, butter, overcoat – have properties of such constant unwavering importance, and have such stereotyped names, that we end by believing that to conceive them in those ways is to conceive them in the only true way. Those arc no truer ways of conceiving them than any others; there are only more frequently serviceable ways to us."
>
> [James, W. (1890) *The Principles of Psychology* (Henry Holt & Co), 222-224]

**1. Introduction**

The ability to make analogies lies at the core of human cognition, intelligence, and creativity (Hummel, Holyoak, 1997: 427). These key abilities are inherently connected to human learning, human reasoning, and human adaptation of knowledge to new domains[3]. The capacity for analogy-making and conceptual abstraction is, therefore, a key feature of human evolution as knowledge is transferred from sources to targets (*e.g.,* from "harm-causing sources to potential harm-causing targets"). Similarity is key to this probabilistic assessment.

In a very simple and basic first approach, the higher the degree of similarity between *x* and *y*, the higher the probability of correctly inferring that *x* has *Q* upon knowing that *y* has *Q* (Goldstone, Son, 2005: 1). Now, this presupposes accepting a given perceived world structure as Rosch put it in 1978: "the perceived world is not an unstructured total set of equiprobable cooccurring attributes". Instead, the material objects of the world are perceived to possess high correlational structure (*i.e.,* it is an empirical fact provided by the perceived world that "wings co-occur with feathers more than with fur") (Rosch, 1978: 29).

Artificial Intelligence ("AI") is the study of cognitive processes with the specific use of conceptual frameworks and tools of computer science (Rissland, 1990: 1958). General themes of cognitive science, such as such as analogy or categorization, among others, are therefore specifically placed in the context of AI. Whether or not the placement of these themes in the

3. On evolution and adaptation, ENDLER (1986: 5).

context of AI alters the general problems depends upon the themes and the problem themselves. This means that while some aspects of analogical reasoning in AI simply have to do with background theories of analogy and their constituents in general, others, in turn, will have to do with the applicability of those theories in the context of AI (or the feasibility thereof).

It is said that, while AI has progressed substantially and exponentially in the last decades, notably in robotics and Natural Language Processing ("NLP"), AI systems still lack the capacities and robust human conceptual knowledge, transversal to different contexts, necessary to make proper analogies. It has proven tremendously difficult to replicate the analogical human mind by a machine. "No current AI system is anywhere close to a capability of forming humanlike abstractions or analogies" (Mitchell, 2021: 79). There are several reasons for these shortcomings. I shall mention a few as I do not intend to exhaust them here.

For starters, AI systems do not "experience" stuff. There is an old Portuguese popular saying: "a scalded cat is afraid of cold water". I am vaguely certain that a cat will eventually be able to assimilate through natural processes that it was the property of "hotness", not that of "liquidness", that caused its painful reaction. Up until then, the cat is simply *wrong* as its concerns are rather overinclusive. If cats are avoiding getting scalded, they should indeed be afraid of hot things, not of liquid ones (unless they are also hot); they should be afraid of liquid things if circumstances are that they might drown, whether or not the water is hot. But it takes time and intelligence to perform this kind of differential diagnosis. Humans assimilate it rather easily.[4]. Machines, on the other hand, cannot assimilate it unless they are reasoning with built-in predefined data. Machines do not *experience* pain.

Take, for instance, the background of an experiment conducted by RISSLAND:

> (…) *knowledge about the problem case is input directly by the user. Knowledge about what makes one case similar to another* – particularly what makes one case a good precedent to appeal to in making a legal argument about another – *is embedded in HYPO-style CBR*. Knowledge of the mechanics of forming a query is handled by the relevance feedback mechanism of INQUERY Knowledge about the domain (e.g., personal bankruptcy law) is used in the CBR module, and knowledge about text (e.g., word frequencies) is used in the IR module.(…) (Rissland, 1995: 401)

The necessity to input and embed knowledge into AI systems triggers the problem of knowledge elicitation bottleneck ("KEB"). On the one hand, the task of eliciting knowledge from domain experts is the most difficult and

4. This is tantamount to Schauer's argument regarding the "dogness" (not the "blackness") of Angus, the black Terrier, causally and relevantly justifying the "no dogs allowed" rule. See SCHAUER (1991: 25).

time-consuming task in developing a knowledge-based system.[5]; on the other, it has been proven quite difficult for anyone to simply "describe" knowledge, let alone encode it in a pure manner into an AI system without conflating description and evaluation (hence the *bias effect* problem in modern AI).

Replacing Rule-based AI systems ("RBS"), that include explicit representation of general domain knowledge, with Case-based AI systems ("CBS"), that merely track memory of solved situations, just partly tackles the KEB problem.[6]. It is perhaps too optimistic to claim, as Liao, Zhang and Mount that Case-based Reasoning ("CBR") not only parallels the way people solve problems in various situations as it also "overcomes the drawbacks of rule-based systems" as it "does not require an explicit domain knowledge model, and the problem of KEB can be avoided" (Liao et al., 1998: 267). To the extent the main implementation of CBS is "reduced to identify significant attributes" and *then* take advantage of problem cases as they arise, the KEB problem is still there in one of its variants. Any expert struggles to easily "describe" how he views a problem and sharply distinguish between facts, beliefs or any other factors that actually influence his decision-making. And even CBS cannot do without certain paramount rules.

Secondly, the fact that analogies are used to achieve many different goals is a problem for AI analogy-making. According to Holyoak, the process of analogy is governed by core constraints provided by isomorphism, similarity of elements, and the goals of the reasoner (Holyoak, 2012: 136). Analogies may be used in explanations, when a source analog is used to provide understanding of a target phenomenon, to form hypotheses and help evaluate them (such as the case in which we hypothesize that a creature with wings will most likely fly) or to present political, historical, and legal arguments, aimed at convincing an audience that a particular conclusion is warranted (Thagard et al., 1990: 260). Indeed, they intrinsically depend upon the goal of the analogist: analogies are *target-oriented.*

"Analogies are used to achieve the goals of the reasoner (...)" and "mapping is guided not only by relational structure and element similarity but also by the goals of the analogist." (Holyoak, 2012: 124)

"One important reason to learn categories is that they provide a basis for inference: knowing that an item belongs to a category allows one to infer many additional characteristics about the item. Another is that categories have goals." (Seger, Peterson, 2013: 17)

5. "Building enough training data for a supervised method of similarity measurement is a costly process, as it requires legal experts to manually analyze the document text and assign a degree of similarity for a given pair of documents". See Mandal et al. (2021: 5).
6. It is still time-consuming to build a case corpus of significant size if cases are represented in any depth. If the case base is constructed after the fact from pre-existing archives of textual materials, the task can be daunting. See Rissland (1990: 401).

Now, it has proven difficult to program AI systems to be functionalized to one purpose (necessarily a *a priori* built-in purpose), let alone the multiple purposes that may abstractly guide different types of analogical reasoning[7]. For instance, if we think about legal automated reasoning (ultimately automated judicial decisions) an AI decision-maker would not simply be analogizing for (re)constructing the rule applicable to a legal gap; analogies would also be necessary for explanatory and evaluative purposes. AI systems need to analogize with different goals that vary in accordance with the tasks at hand. It suffices to think about either analogies between causal processes and the evidentiary reasoning underlying a trial or, on the other hand, the analogies required for purposive interpretation of legislative data. But even solely in (re)constructing the rule applicable to a legal gap, it seems rather clear that an AI system would be *expanding a category or classification* (one which includes both the target and the source). For that, the AI system should be able to provide an explanation to substantiate categorization and to include some kind of criterion for relevant and sufficient similarity. And that is one tough cookie to grasp with.

"Announcing an answer to a problem without an explanation, even if it is based on the best practices of the inference technique involved, isn't enough. At present, while a host of methods can perform concept learning and classification, for instance, fewer can provide explanations, say, via analogies with past exemplars." (Rissland, 2006: 40)

Thirdly, AI systems are currently uncapable of successfully transferring memory data (*i.e.,* what has been learned) to cases and situations outside their training regimes. Analogy entails flexibly mapping familiar relations from one domain of experience to another and "classical or symbolic AI models lack the flexibility to apply predicates or operations across diverse domains" (Hill et al., 2019: 1). AI systems inevitably fall more or less prey to adversarial examples or change of context. It is not simply a problem of monotonic versus non-monotonic automated reasoning; machines are currently unable to understand extended and metaphorical notions. Borrowing Mitchell's example, AI systems have a difficult time linking notions such as "water bridges," "ant bridges," "bridging one's fingers," "the bridge of a song," "bridging the gender gap," "a bridge loan," "burning one's bridges," "water under the bridge," and so on. In legal parlance, AI systems may have a hard time applying the concept of "good faith" to different contexts such as "*acting* in good faith" (a normative standard) or "*being* in good faith" (a psychological status). Rather unhumanlike errors are therefore bound to happen.

7 The work of KEDAR-CABELLI (1988) takes a limited step in this direction by employing a notion of 'purpose' to direct the selection of relevant information, but still starts with all representations pre-built. (CHALMERS ET AL., 1992: 198). General heuristic "means-end analysis" has been applied in AI by means of analyzing differences between the goal state and the current state of a problem and to strategically apply operators to reduce the difference. See DE CORTE ET AL. (2012:1423).

Finally – though I have no intention to be exhaustive –, there is no sufficient consensus over theoretical background theories of AI analogy. There are several general models of human analogical reasoning, precisely the reality that AI analogy-making intends to mirror. This paper intends to expound on some theoretical aspects of these shortcomings. It will focus specifically on similarity, a necessary yet insufficient step in warranted analogical conclusions and categorization by AI Systems.

## 2. Conceptual apparatus of analogy: non-identity, similarity, relevance and equality

I will now briefly outline some of the definitions used below: "identity", "comparing", "similarity", "relevance", "category" (and "class"), "equality" and "partitioning". I will follow closely the definitions I have laid down elsewhere as they merely serve the purpose of clarifying subsequent arguments in this paper (Lopes, 2022: 331-382).

"Identity": I take "identity" to mean that *two or more objects have in common all their properties;* they are therefore but one single object (or term). This is so under the Leibniz law of the identity of indiscernibles, formalized as ($\forall x \, \forall y \, \forall \varphi \, ((\varphi x \rightarrow \varphi y) \Leftrightarrow x=y)$).[8].

"Comparing": I take "comparing" to be a *specific mental action which entails examining or looking for similarities and dissimilarities between two or more objects or terms*; it requires two objects, instruments of measurement and probably – although this is not consensual – commensurability (or, at least, comparability).[9].

"Similarity": I take "similar" to mean that *two or more objects (particulars) [or terms] are in a relation such that they are perceived as sharing (or resemble sharing) one or more properties*.[10]. This is the very basic notion of similarity, consisting of relations between "particulars" and "properties":

> In asserting a "similarity fact" between two or more objects or terms, a given property (or more) is/are required to be singled out: *a* and *b* are always (dis)similar with regards to a given property. Asserting that "*a* and *b* are similar" is an elliptical statement (and so is asserting that "*a* and *b* are equal").
>
> Assertion of "similarity facts" merely commits the utterer to two inferences (Cumpa).[11]:

8. If placement of objects in spatial context is seen as a relevant property, then no identity can ever be assigned to two different objects. See COMANDUCCI (2010: 32). On the principle of identity (as applied to propositions), see ECHAVE ET AL. (2008: 83 ff.).

9. Do note that comparing does not entail "treating equals equally" and vice-versa.

10. According to RUSSELL, a relation unites terms: "a relation is distinguished as dual, triple, quadruple, etc., or dyadic, triadic, tetradic, etc., according to the number of terms which it unites in the simplest complexes in which it occurs." (1911: 5).

11. The assertion of "similarity facts" ontologically commits oneself to the existence of similarity relations, but not that such similarity relations must be universal or of other metaphysical nature. (CUMPA, 2021: 8-9).

> (a) a "relational inference" by which "if a stands in similarity relation Sq to b, and c stands in similarity relation Sq to d, then there is a similarity relation Sq with four terms" (and other terms may be added);
>
> (b) a "neutralist inference" by which "if a stands in similarity relation Sq to b, then there is a similarity relation, Sq (a, b)".

"Relevance": I take "relevance" to be *a predicate of a property of a particular, that which denotes the relation of a property of a particular with a given standard* (*e.g.*, a legal, moral or natural system).[12]. A property of a particular is said to be relevant if and only if, when contrasted with a given standard of relevance, the property of that particular instantiates a hypothetical property represented in that standard. The assertion of "relevant similarity facts" between two objects thus presupposes *something more* than just the mere assertion of a similarity fact but still *something less* than categorizing.

"Category": I take "categories" (or "classes") to be *collections of sets that can be unambiguously defined by a property that all its members are represented to "share" (or resemble sharing)*.[13].

"Equality": I take "equal" to mean that *two or more objects (or terms) are said to belong to the same class or category, since they share "relevant" properties according to a given standard of evaluation*. Judgments of equality entail selecting a standard of evaluation of two or more objects which, in turn, amounts to highlighting one or more relevant properties that preside over the comparison and determine the categorization.

> Equality is not described, rather *it is declared*. Further to acknowledging relevant similarities, equality *also* entails disregarding (irrelevant) dissimilarities between two particulars, much like constructing a self-constricting "tunnel-like" field of vision that only considers properties which are perceived to be shared.[14].
>
> The illocutionary force of "setting up", "establishing" or "instituting" is not assertive, rather declarative: "declaring" entails a performative effect. In creating the category with members *a, b* and *c* I perceive as a result *a, b* and *c* not just as particulars (*e.g.*, dog, rabbit and whale), or even as particulars with a similarity relation with regards to property q (*e.g.,* animals with milk-producing mammary glands). Quite differently, I perceive them, and they present themselves to me as "particular x's, being

---

12. On "relevance" as a relation, claiming that A may be said to be relevant only in respect to a given B, see YOVEL (2003: 309). See also RATTI (2013: 61).

13. On the several meanings of "category", see STRAWSON (1974: 25). If one endorses philosophical nominalism, then only singular things really exist: species, sets, classes, categories, common natures, general properties, shared attributes, and so on, are thus taken to have no mind-independent reality of their own.

14. *A* and *b* are *judged*, *evaluated, established, instituted,* or *declared* equal; they are not simply *described* as similar.

instances or tokens of more encompassing categories" (mammals) (Schauer, 1991: 18). Naturally, this is conditional on the individual declaring having an active disposition – *i.e.,* a certain type of power – to do just that (Searle, 1976: 13).

The declaration of equality is a type of decision-making. It has been stated by neuroscientists that definitions of "decision-making" and "categorization" are similar up to the point that only categorization works on generalizable representations. In fact, generalization is said to be the key distinction between decision-making and categorization (Seger, Peterson 2013: 1189). For instance, lawmakers do not describe nor prescribe relevant properties: they rather prescribe actions and decide to *"declare" relevant properties*. However, when enacting a legal provision, the lawmaker is not prescribing relevance; she is not stating that property *x* ought to be relevant; rather she is stating that property *x* "is" relevant, therefore legal consequences apply whenever property *x* arises in a case.

"Partitioning": a judgment of equality entails partitioning the set of objects that can be evaluated under such standard into two mutually excluding and jointly exhaustive classes:

(a) the class of objects equal amongst themselves under a given standard of evaluation and;

(b) the class of all remaining objects.

A judgment of difference, on the other hand, entails a partition of the set of objects that can be evaluated under such standard in two mutually excluding and jointly exhaustive classes:

(a) the class of objects different amongst themselves under a given standard of evaluation

(b) the class of all remaining objects.

Lawmakers "construct" standards of relevance. "To equate different things is the real task of the legislator" (Gianformaggio, 1997: 271).[15]. But in equating different things, the lawmaker necessarily differentiates. From a set of particulars (*a, b, c, d... n*) the lawmaker establishes an equivalence between (*a, b, c*), thus forming "category x", the members of which are *a, b, c*. In doing so, the formation of this category necessarily leaves out particulars *d* to *n*. "Equation" between a, b, c occurs in forming "category x" but "differentiation" occurs between members of "category x" and members of "category non-x". Addressees, actions and state of affairs may be *(i) inside* the standard of relevance several "classes of addresses", "classes of hypothetical action-types" and "classes of occasions", those which are *equal under that standard of relevance* and *(ii) outside* the standard of relevance several "classes of non-addresses", "classes of non-hypothetical action-

15. "The law must (...) refer to *classes of persons*, and to *classes of acts, things and circumstances*" (HART 1994, 128).

types" and "classes of non-occasions", those which are *different under that standard of relevance*.

In establishing the equivalence, declared equality presupposes creating a rule – a "standard of evaluation", "meta-factor" or "criterion" (I have called it "RuleCatg"). It does not merely set out the relevant properties under comparison and the discarding of all the other (irrelevant) properties. It prescribes a "thesis of relevance" and a "thesis of irrelevance". The RuleCatg functions as an exclusionary reason that unburdens the decision-maker in comparison and class-creation.[16].

Having clarified the meaning of some terms to be used below, I now move to the inferential model of analogy-making.

## 3. Inferential steps in the analogy: narrowing down to "relevant" and "sufficient" similarity

### (i) Inferential steps in general analogies

Analogy presupposes comparison between two non-identical objects or terms. From the perspective of cognitive science, according to HOLYOAK, analogical reasoning is "the complex process of retrieving structured knowledge from long-term memory, representing and manipulating role-filler binding in working memory, performing self-supervised learning to form new inferences, and finding structured intersections between analogs to form new abstract schemas.[17]".

From the argumentative point of view, the argument from analogy (or *a simile* argument) is the process of inferring that a "property-conclusion" *Q* ought to be predicated to a particular situation or object *T* ("target") from the fact that *T* shares a property (or set of properties) *P* with another situation / object *S* ("source") which *has* property *Q*. The set of common properties *P* is the similarity between *S* and *T* and the "property-conclusion" *Q* is projected from *S* onto *T* (Davies, 1988: 228-229). See the following schema:

P(S) A Q(S)

P(T)

___________

Q(T).

As seen above, in AI, RBS retrieve structured knowledge from the embedded data of the legal system whereas CBS retrieve structured knowledge from past cases and decisions. For the latter purpose, I am assuming that a case is a schematic description of a problem by an index (by means of n-tuples)

16. On rules as "exclusionary reasons" see RAZ (1999: 35 ff.).
17. See HOLYOAK (2012: 136).

of the main characteristics of each element of the domain problem (Rodriguez, 2001: 133).

Example from memory case

| | |
|---|---|
| Applicant age | 30 |
| Grant amount | 1000 |
| Research purpose | Law & AI |
| Job occupation | Professor |
| Job contract type | Tenure |
| Solution: monthly payment | 50 |

There are two main problems in analogical reasoning: the (i) non-redundancy problem and the (ii) problem of justification (Davies, 1988: 229).

The non-redundancy problem is connected to the insufficiency of the "database", *i.e.,* it has to do with its extension of the database and the fact that analogy is conceptually an inferential discovery. In analogical reasoning, the background knowledge that requires an analogy must be *per se* insufficient in providing the conclusion (the "target" of the analogy shall not be able to solely provide for the solution, *i.e.,* it cannot be identical to the source); symmetrically, the "source" must necessarily provide information which is not contained in the "database". If one has complete knowledge about the reasons why an object has a property, then general similarity is no longer relevant to generalizations (Goldstone, Son, 2005: 14). The problem of justification has to do with finding a sound criterion which, if met by any particular analogical inference, sufficiently establishes the truth of the inference; it has to do with the validity or soundness of the inference and the fact that analogy is conceptually an inferential justification.

While the problem of justification seems to be more of a general problem of analogical reasoning – notably in legal analogy in which verification cannot be carried out –, the non-redundancy problem poses specific questions in AI reasoning. For proper AI analogy to take place, AI systems would necessarily (i) require analogical reasoning *vis-à-vis* a target absent prior encoded complete knowledge about the reasons why that target has a given property and (ii) not having been fed with encoded knowledge about that target. If this were the case, this would probably mean an AI System passing the Turing test.

Taking into consideration the above, one can state that analogy in general is a complex inference that includes three inferential steps (Canale, Tuzet, 2009: 499-509).[18].

| | Abduction (step 1) | Induction (step 2) | Deduction (step 3) |
|---|---|---|---|
| Rule "all *x* that is *P* is *Q*" | Premise | Conclusion | Premise |
| Case: *x* is *P* | Conclusion | Premise | Premise |
| Conclusion: *x* is *Q* | Premise | Premise | Conclusion |

Assume, as the source case, that S is *P* and S is *Q.* The steps are the following:

1. an abduction of relevant property *P*:

property *P* (that one knows S has) is *the best available explanation* for property *Q* (that which one knows S has and one wants to know whether T also has)].

2. an induction of the class of members that have such property *P*:

P1: *S* that is *P* is *Q*

P2: *S'* that is *P* is *Q*

P3: *S''* that is *P* is *Q*

C. "all x that is *P* is *Q*"

3. a deduction of the target having property *Q*:

P1. "all x that is *P* is *Q*"

P2. T is *P*

C. T is *Q*

Whether or not the analogy is sound depends on whether the explanatory theory is soundly shared between source and target, the inferences are valid, and the transferred knowledge holds. As James put it in the quotation at the beginning of this paper, "paper" is compared to "canvas" as a means to obtain the valid inference of paper having the dispositional property

18. I am using the table of J. Maranhão (2012: 83 ff.).

"writable" because James wanted to write. James paired "paper" and "canvas" *vis-à-vis* the dispositional passive property "writable"; if James wanted to light a fire, "paper" would have been compared to "wood" as a means to obtain the valid inference of paper having the dispositional passive property of "combustible". James would have paired "paper" and "wood" *vis-à-vis* the dispositional property "combustible".

(ii) Inferential steps in normative analogies

In normative inferences, the structure of the analogical argument, according to Peczenik, goes like this.[19]:

> P1: if the fact F or any other fact, relevantly similar to F occurs, then Q is obligatory ("meta-level premise", connected to the relevance of similarity);
>
> P2: H is relevantly similar to F ("object-level premise", connected to the case);
>
> C: if H then G is obligatory.

Although PECZENIK's syllogism adds up to ALEXY's syllogism – which did not differentiate between "relevant similarity" and "irrelevant similarity" – it does not *differentiate clearly* between "relevant similarity" and "sufficient similarity"; perhaps PECZENIK includes "sufficient similarity" in "relevant similarity").

In PECZENIK's syllogism, the abductive step is encapsulated at the metalevel of "relevant similarity". It should, however, be stressed that the abductive step of providing for an explanatory theory has nothing to do with similarity *per se*; it may have to do with "relevantly (similar)", though, but that highly depends on the meaning ascribed to "relevantly". Consider a man jumping into a swimming pool fully clothed, in the example provided by MURPHY and MEDIN. This man may be categorized as "drunk" to the extent one comes up with a sound theory of behavior and inebriation that consiliently explains the man's action, but *not because there is similarity between the instance and the category*.[20]. I am referring to "consilience" as explanatory power, including how much a hypothesis covers, how fruitful a hypothesis is in suggesting interpretations of observations, and the connections a hypothesis establishes between various observations (Schvaneveldt, 2012: 14).

---

19. See PECZENIK (1989: 321); ARASZKIEWICZ (2011: 102).
20. "[T]he categorization of the man's behavior does not depend on matching the man's features to the category drunk's features. It is highly unlikely that the category drunk would have such a specific feature as "jumps into pools fully clothed." It is not the similarity between the instance and the category that determines the instance's classification; it is the fact that our category provides a theory that explains the behavior." See MURPHY, MEDIN (1985: 295).

Peczenik's syllogism may be paralleled to the previous schema in the following manner:

Source case: S is *P*; assuming that $[S_1, \ldots S_n] \Longrightarrow^n Q$ we obtain that $[S_1, \ldots S_n] \subset Q$. By accepting that the application of a rule R to a given case C consists in adding the outcome of the rule to the set of facts of the case, we obtain that S is *P* and S is *Q* (the latter *qua* outcome of rule R applicable to S].[21].

1. an abduction of relevant property *P*:

property *P* (that one knows S has) is the best available explanation for $[S_1, \ldots S_n] \subset Q$ (that which one knows S has and one wants to know whether T has in order to know if $[T_1, \ldots T_n] \subset Q$) holds]

2. an induction of the class of members that have such property *P*:

P1: *S* is *P* and S ⊂ *Q*

P2: *S'* is *P* and *S'*⊂ *Q*

P3: *S''* is *P* and *S''*⊂ *Q*

C. "all x that is *P* is *Q*"

3. a deduction of the target having property *Q*:

P1. "all x that is *P* is *Q*"

P2. T is *P*

C. T is *Q*

Much like analogy in general, legal analogy is a means to an end. One compares cases (analogs) and retrieves structured knowledge from the legal system or precedents (*means*) in order to form new inferences and find structured intersections between them to form new abstract normative schemas (*end*). One can say that legal analogy is the process by means of which:

(i) one wishes to establish whether a certain $case_1$ that is *P* and also is *Q* while aiming at a certain purpose *(end)*;

(ii) for that purpose, one compares $case_1$ (T) with *non-identical* and previously established as *relevantly similar* $case_2$ (S), that which one knows is *P* and *Q* (*means*);

21. On this, Araszkiewicz (2011: 103).

(iii) one comes up with a theory that explains that property *P* is the best possible explanation for *Q* applying to $case_2$;

(iv) a parallel is drawn between *(a)* the objects of the world being perceived to possess high correlational structure (Rosch) and *(b)* the cases governed by a legal system also being perceived to possess high correlational structure as a legal system shall not be an unstructured total set of equiprobable cooccurring deontic statuses].[22];

(v) one establishes that $case_1$ is *sufficiently similar* to $case_2$ (*i.e.,* similar in a degree that is sufficient to justify that $case_1$ is declared *equal* to $case_2$ and the outcome of the rule ($\subset$ Q) is transferred from S to T: $[T_1, \ldots T_n]$ ($\subset$ Q).

(iii) Verification-apt and non-verification-apt analogies

Some analogies make room for empirical testing of the inference while others do not. The first step of analogy is abductive reasoning: coming up with a suitable theory that consists in applying the norms underlying the generation of hypotheses (Schvaneveldt, 2012: 12). The establishment of whether the object "paper" has the dispositional passive property of "combustible" (with the aim of lighting a fire) can be verified by lighting a match close to the object and observing whether it burns (*i.e.,* whether *Q* obtains for T).

In the realm of normative inferences things are rather different.

Let us recall SCHAUER's case of the Angus, the black Terrier. When formulating the norm statement "no dogs allowed in the restaurant," the restaurant owner availed himself of the background knowledge regarding the previous disarray imputed to the "dogness" of Angus, a black Scottish Terrier, and generalized from Angus to the class of "dogs" of which Angus is a member.[23]. It is important to stress that the outcome "disarray" is not imputed to Angus but to the property of "dogness" (and not that of "blackness" or any other property of Angus). This is a theory: it correlates the causally relevant property of "dogness" (predicated to Angus but also to the entire class of dogs) with the active disposition to create disarray.

Let us now break the former sentence into components and ascribe the steps of the inferential process:

(i) "the restaurant owner availed himself of the background knowledge" (*retrieval of structure knowledge*);

22. Usually, legal systems include rules governing analogies. Nothing prevents that the system is incoherent, though.
23. And in doing so, the lawmaker is presupposing that it is the property of "dogness"—and not, for instance, that of "blackness"—that was cause to the consequence of the disarray at the restaurant. See (SCHAUER 1991: 25).

(ii) "(…) regarding the disarray imputed to the [causally relevant property of] "dogness" [of Angus]" (*abduction* or *best explanatory theory*);

(iii) "(…) of Angus, a black Scottish Terrier" (*source*);

(iv) "(…) and generalized from Angus" (*induction*);

(v) "(…) to the class of "dogs"" (*set of targets* categorized);

(vi) "(…)of which Angus is a member" (*deduction* from the category).

Setting aside the naturalistic fallacy, it is fairly reasonable to conclude that the owner should enact a defeasible norm according to which "all people with dogs ought not to enter the restaurant". But reasonings such as this one pose several problems. I will point out three.

The first problem has to do with the fact that "causality" does not apply in normative inferences. In fact, even "causal relevance" can only seldom be used in warranting analogies. Striking an analogy between Angus and other dogs with the aim of governing the action of "entering the restaurant" as forbidden works persuasively in SCHAUER's example. But striking an analogy between heterosexual couples and homosexual couples with the aim of governing the action of "marrying" as permissible does not make room for empirical evidence. It is not a causal scenario such as lighting a fire with a paper. One cannot empirically verify whether *Q* (as the outcome of the rule $[T_1, \ldots T_n] \subset Q$ (homosexual couples) analogous to the source-rule $[S_1, \ldots S_n] \subset Q$ (heterosexual couples) holds in *T*. In other words, one cannot empirically verify whether there is relevant and sufficient similarity between source and target analogs.

In testable causal scenarios we can make use of "causal relevance" as the underlying justification, according to Schauer, of the *prescription of the goal*, or *proscription of the evil*, that the lawmakers intend to procure by relying on relevant background knowledge over previous particular cases. Underlying justifications are "relevant reasons for φ". For instance, the fact that the object "paper" has the dispositional passive property of "combustible" is an explanation of why I did use paper with the aim of lighting a fire but also a reason for me having done so. But in law, relevant reasons do not refer to explanations of causality, rather they refer to the much confusing and criticizable notion of *ratio legis* (a shady way to sabotage the law, according to TARELLO). As CANALE and TUZET put it, "it is the ratio which establishes what is relevant for what. But, as legal scholars know well, the determination of the ratio is often a difficult and controversial task" (Canale, Tuzet, 2009: 499). It is in view of this that Canale, Tuzet attempt at using Brandom's "scorekeeping practice" in claiming that the relevant reasons are determined by the normative statuses reciprocally attributed by the speakers in the context of legal argumentation. But if this is hard for humans, all the more for AI systems.

The other two listed problems are specific to AI. They are connected (again) to the KEB.

On the one hand, the second problem is that AI Systems (either RBS or CBS) must be subject to external data input as to whether legal reasonings are defeasible and permit analogy or, in turn, whether *a contrario* applies.[24]. Even CBS must be fed with closure rules that allow for *a contrario*. But if this is easier for *nulla poena sine lege*, it becomes more complicated with the embedding of parameters for exceptionality that presuppose an input on all general norms *vis-à-vis* which a norm is an exception.

On the other hand, the third problem is that AI Systems face a possible paradox: the analog cannot be something to which the database already has an answer to (in which case it is not an "analog"). The underlying justification of a normative outcome can be structured as a rule-justifying principle: this is common knowledge in the literature that acknowledges the differences between rules and principles or, at least, that rules are enacted for the purpose of pursuing goals (it matters not, now, whether goals ought to be framed as "principles"). But here is the gist: if the AI System is previously subject to input by the expert with the applicable principles and goals, there is no analogy *per se*, simply the application of the principle or goal to the instantiating case. We are left with AI systems that can move from principles and goals to decisions. Irrespective of whether purpose comes into play, this is not much, is it? The AI system is not really performing an equality judgment so much as it is merely acknowledging that both "source" and "target" are members of the set of facts that instantiates the principle embedded in the database. Now, to the extent there is no "gap", there really is no "target" here, is there?

## 4. Models of similarity measurement

The theoretical framing of similarity (a necessary yet insufficient condition of analogy.[25]) is divided at least into three accounts: *(i)* geometrical models; *(ii)* featural models and *(iii)* alignment-based models. I will mostly focus on the featural model, but I will briefly address the other two as it becomes relevant.

### (i) Feature-based models of similarity

In his seminal paper "Features of Similarity", AMOS TVERSKY described similarity-relations in terms of a feature-matching process based on weighting common and distinctive features between objects in comparison (Tversky, 2007: 327 ff.). TVERSKY designed a "contrast model" aimed at criticizing the shortcomings of the mainstream geometrical models of similarity (Shepard, 1962) that asserted similarity relations between entities in which the degree of similarity between a pair of objects was taken to be inversely related to the distance between two objects' points in a dimensionally organized metric space: $sim\ (A,B) = \varphi\ (dist\ (A,B))$ where *dist*

24. On this, ARASZKIEWICZ (2011: 104).
25. Claiming that analogy is a special kind of similarity, see HOLYOAK (2012: 117).

is the relation of distance having values from 0 to 1; $\varphi$ is a decreasing function satisfying $\varphi$ (0) = 1.[26].

Tversky conducted several experiments challenging the three axioms of the geometrical model:

(i) Axiom 1 – Minimality: similarity between an object and itself is the same for all objects.

*Counterexample*: if identification probability is interpreted as a measure of similarity, then conclusions that an object is identified as another object more frequently than it is identified as itself violate minimality.

(ii) Axiom 2 – Symmetry: distance from A to B should be the same distance or equal to the distance from B to A.

*Counterexample*: logical symmetry may be true but similarity assertions are directional (a is similar *to* b) and they differentiate a subject (a) and a referent (b): North Korea (subject) is more similar to China (referent, prototype) than China is similar to North Korea.[27].

(iii) Axiom 3 – Triangle Inequality: the direct path from *a* to *c* will not exceed the indirect path through *b*, and if *a* is quite similar to *b* and *b* is quite similar to *c*, then *a* and *c* cannot be very dissimilar from each other.

*Counterexample*: Jamaica is similar to Cuba (because of geographical proximity); Cuba is similar to China (because of their political affinity); but Jamaica and China are not similar at all ("similarity, as one might expect, is not transitive").[28].

---

26. See IFENTHALER (2012: 2147). In Multidimensional Scaling, the distance between points *I* and *j* is computed by a complex formula:

$$dissimilarity(i,\ j) = \left[\sum_{k=1}^{n} |X_{ik} - X_{jk}|^r\right]^{\frac{1}{r}}$$

Where *n* is the number of dimensions, *Xik* is the value of dimension *k* for item *I*, and *r* is a parameter that allows different spatial metrics to be used. With r = 2, a standard Euclidean notion of distance is invoked, whereby the distance between two points is the length of the straight line connecting the points. I am quoting from GOLDSTONE / SON, Similarity, 15.
27. Differentiating between logical symmetry and epistemological asymmetry in similarity assessments, on a remark to ADLER (2007: 83–92), see CANALE, TUZET (2009: 500, note 2).
28. Tversky concluded that "minimality is somewhat problematic, symmetry is apparently false, and the triangle inequality is hardly compelling." See TVERSKY (1977: 329).

In TVERSKY's contrast model, similarity is a function of the number of features or prominent attributes (weighed according to its relevance or salience) perceived as being shared and not shared between objects in comparison. Assertions of similarity are feature-matching processes measured as a linear combination of their common and distinctive features. Similarity is thus seen more as a function that measures the degree to which two sets of features match each other, rather than as a metric distance between points in a coordinate space. It increases through common features and decreases through unique features of the objects. A feature may be any property, characteristic, or aspect of a stimulus: there are no limitations to what it is. Features range from "surface attributes" such as color or shape to less accessible, deeper attributes as "hidden causal powers" or "dispositions" (Komatsu 1992: 514).

The computation of similarity in the contrast model abides by the following formula, which will provide for the weighted sum of the measures of all of the common features within a category minus the sum of the measures of all the distinctive features:

$$S(A, B) = \theta f(A \cap B) - af(A - B) - bf(B - A)$$

(i) $f(A \cap B)$ denotes common features of objects A and B;

(ii) $f(A - B)$ denotes features exclusive to object A;

(iii) (B – A) denotes features exclusive to object B. $\theta$, $a$, and $b$ are weights (constants) for the common and distinctive components that may vary in accordance to underlying theoretical assumptions of relevance and salience.

The contrast model allows for asymmetric similarity because $f(A - B)$ may not equal $f(B - A)$. Moreover, any dimensional scaling can be translated into featural representations: dimensions, as factors, are gradual (bright is more light than dim, and large is more size than small), therefore, according to Tversky and Gatti, the features of B include a subset of A's features whenever B is brighter, or larger than A (Tversky, Gatti, 1982: 123 ff.).[29].

The contrast model can be transposed to AI CBS by assessing common characteristics (features) of cases in two ways: global similarity and local similarity.

The local similarity is evaluated when the same feature is compared in two different cases and returning values are either 1 (True) or 0 (False). Then, the global similarity of cases is easily obtained as it combines the many local similarities into a single measure. Using Rodriguez's expression, if a case $C_i$

29. On dimensions and factors, ARASZKIEWICZ (2011: 101).

is described by a set of descriptors $\{a_1, a_2, ..., a_n\}$, the global similarity with another case $C_j$ is given by:

$$SF\ (C_i, C_j) = F[Sf_1(a_1, b_1),\ Sf_2(a_2, b_2),\ Sf_3(a_3, b_3),\ ...\ Sf_n(a_n, b_n)$$

Where $Sf_n$ are the functions evaluating the local similarity between the two cases' descriptors and returning values between 0 and 1.

Local similarity function will return "True or 1" when the two elements have the same logical value and "False or 0" in any other case (Rodriguez, 2001: 137). Local similarities are more difficult to assess when numeric values are compared as they presuppose a complicated mathematical exercise of setting a relevant window over the range of values for the numeric attribute.[30]. If the attribute value for $C_j$ is outside the relevant window, the similarity between the two values is null and the $Sf_n$ function will return 0. If the values are identical, the $Sf_n$ function will return 1 (maximum). In any other case, the $Sf_n$ function will return with a value between 0 and 1, which varies according to the proximity of the values.

(ii) Directionality and diagnosticity in similarity assertions

The empirical experiments of Tversky were conducted from the perspective of cognitive psychology but they give much insight into similarity relations in analogical reasoning and AI. The main tenets are "directionality" and "diagnosticity".

Relevant similarity relations are "directional" and the choice of subject and referent ("*a* is similar to *b*" or "*b* is similar to *a*") depends much on the relative salience or prominence of the objects. Tversky claims that the variant is more similar to the prototype than the prototype is to the variant, because the prototype is generally more salient than the variant (Tversky, 1977: 333). If one is evaluating whether *a* is similar to *b*, then *a* is the subject of the comparison and *b* is the referent (the "prototype"). In such a task, one naturally focuses on the subject of the comparison. The contrast model accommodates non-mirroring between similarity and dissimilarity judgments.

In analogical reasoning, the roles of the "subject" and the "referent" are played by the "target" and the "source". In analogizing, one compares the target to a potential source and not the other way around. Epistemologically, one naturally focuses on the target. It is not a matter of salience but a matter of the referent being accounted for in the system (and that is usually the case with the "prototype"). Do note that this directionality is linked simply with the fact that one is asserting similarity facts between the target and the source. The common features term ($A \cap B$) is hypothesized to receive more weight in similarity than dissimilarity assertions; but if one is asserting dissimilarity facts – such as one is performing a *distinguishing* from

30. For more on this, see RODRIGUEZ (2001: 137-138).

precedents – the distinctive features term receives relatively more weight in dissimilarity judgments. The relevant question for AI is to determine whenever an AI System falls prey to this "enlargement of weight" of similarities in similarity assertions as a consequence of the human data input by the expert (or whether directionality is simply a cognitive bias of some sort to which AI Systems are immune).[31].

The diagnosticity principle, on the other hand, refers to the classificatory significance of features, that is, the importance or prevalence of the classifications that are based on these features. According to it, adding or subtracting objects can alter the categorization of the remaining objects. Any change in grouping changes the similarity relations intra-category. TVERSKY provides evidence for the diagnosticity hypothesis with an experiment in which several subjects asserted that Sweden ($subject_1$) was more similar to Austria (referent) than Poland ($subject_2$) and Hungary ($subject_3$) but replacing Poland ($subject_2$) with Norway ($subject_{2'}$) altered the result: the similarity between Sweden and Norway largely increased the number of subjects that asserted that Hungary ($subject_3$) was more similar to Austria (referent) than Sweden and Norway.

In AI, this means that similarity assertion between the target and the source will depend not only on the similarity relations between themselves but also on the similarity relations between $source_1$, $source_2$, $source_3$ and $source_n$. It will largely depend, not only on the gradual extension of the terms of comparison embedded in the database, but also on the similarity relations between those terms of comparison themselves (and the clustering thereof). It thus follows that the inclusion of a new source in the AI database may affect previous analogies made by an AI System.

## 5. Similarity, relevant similarity and sufficient similarity

How is similarity measured as relevant and, or, sufficient? Conceptually speaking, it would appear that sufficient similarity entails relevant similarity but not vice-versa: we can conceive of relevant similarity between two objects or cases but such similarity not being sufficient to, all things considered, include both in the same category. Not the other way around. Additionally, it seems necessary to have both a "standard of relevance" and a "standard of sufficiency": stating simply that *a is relevantly similar to b* and that *a is sufficiently similar to b* is an elliptical statement.

### (i) Relevant similarity

Let us begin with relevant similarity.

TVERSKY's contrast model does not (at least directly) account for the "role" certain features play within their entities. A more accurate similarity model

31. Apparently, they are not immune. "[d]ata coming from AI seems to be intrinsically biased. The reason underlying this issue lies in a very basic – but often underestimated – fact about text corpora on which these algorithms are trained: they are the result of human operations." See MARINUCCI ET AL., (2022: 2).

should, therefore, stem from the featural model but the matching thereof ought to also include "alignment" and "correspondence" between analogs.

In aligned-based models, comparison goes beyond matching features to determine how elements correspond to, or align with, one another. "Matching features are aligned to the extent that they play similar roles within their entities" (Goldstone, Son, 2005: 25). Aligned-based models narrow down the scope of relevant similarities. Consider the wheel of a ship and the wheel of a bicycle. If alignment is not a criterion, the ship and the bicycle share the feature "wheel" and this increases similarity in the featural-based model. But if *alignment is a criterion*, then this matching feature will not increase similarity because the wheel of the ship does not correspond to the wheel of the bicycle to the extent the role they play is different. The similarity is all things considered *irrelevant*. This is enough to show that it is not instrumentally irrational to categorize as equals objects *a* and *b*, and to leave *c* out, even when *a* and *c* are more similar than *a* and *b*.

The demarcation of relevant similarities from irrelevant similarities is the first stage: then we can envisage a degree of relevance with the role played by the weight functions for the common and distinctive components in TVERSKY's contrast model: the $\theta$, $a$, and $b$ in [$S$ (A, B) = $\theta f$ (A ∩ B) − $af$ (A − B) − $bf$ (B − A)].

Alignment can be fed into the AI System through transformational models. Shimon Ullman (1996) argued that in AI objects are recognized by being aligned with memorized pictorial descriptions fed into the system. The candidate objects selected from the comparison is the best match pursuant to an alignment between the subject and all candidate models (Goldstone, Son, 2005: 26). But, again, all this is fed into the system. Despite AI's advances, we are still far from the necessary insight and techniques for the role played by human perceptions in assertions of similarity (and, therefore, in analogical reasoning) to be mimicked in AI systems as we still lack methods for knowledge representation and techniques for dealing with concepts, context, and purpose (Rissland, 2006: 40).

In 1992, CHALMERS, FRENCH, HOFSTADTER claimed that Structure Mapping Engine (SME), an implementation of Dedre Gentner's Structure Mapping Theory (SMT) in AI, simply assumed that perceptions could be represented as collections of attributes and structured relations.[32]. Although the theory expanded from the mapping of properties of analogs to the mapping of structural relations between analogs, CHALMERS ET AL. claimed that knowledge representation for the inputs was based on a notation which was *already* presented in the style of predicate calculus. AI experiments were, in its vast majority, being conducted on the premises that representations were simply structures in predicate logic assuming a "representation module" that supplied ready-made representations (Chalmers et Al., 1992: 185).

32. SMT sets up a strict division between two means of comparing domains of experience; similarity and analogy. "Two domains are similar if they share many attributes (i.e. properties that can be expressed with a one-place predicate like BLUE(sea)), whereas they are analogous if they share few attributes but many relations (i.e. properties expressed by many-place predicates like BENDS-AROUND (sea, solid-objects)". See HILL ET AL. (2019: 2).

The core of the problem, as identified by Chalmers at al., was that AI research bypassed perceptual processes central to analogy in making sense of complex data at an abstract conceptual level, specifically that of "high-level perception" (Chalmers et al., 1992: 186; 191). Mapping of properties and structural relations between source and target was, for many years, taken to be the sole central aspect of analogy which, in turn, was based on predefined representations built in the system. But mapping properties and structural relations between predefined representations is half the job, perhaps the easier one (although extremely hard!), in analogy-making. If the matching is made between built-in structures of predicate logic that were designed specifically (or unconsciously) to allow for said matching, we are simply navigating the tip of the iceberg. What leads up to those structures of predicate logic is left below water.

Contrary to the SMT, the High-Level Perception (HLP) theory of analogy (Chalmers et al., 1992) frames analogical reasoning as a function of tightly interacting perceptual and reasoning processes. In the HLP model, the creation of stimulus representations and the alignment of those representations are *mutually dependent*.[33]. CHALMERS ET AL. distinguish between:

*(i)* low-level perception, involving the early processing of information from the various sensory modalities (akin to Kant's faculty of Sensibility); and

*(ii)* high-level perception, which involves extracting meaning from the low-level perceived data by accessing concepts and making sense of situations at a conceptual level.

This latter specific type of perception ranges "from the recognition of objects to the grasping of abstract relations, and on to understanding entire situations as coherent wholes".[34]. But "high-level perception" mechanisms presuppose AI "low-level perception" mechanisms which, in turn, are far from settled. Do note that CHALMERS ET AL. made this remark in 1992. 30 years later, it is still clear that "no current AI system is anywhere close to a capability of forming humanlike abstractions or analogies" (Mitchell, 2021: 79).

It would be unfair to challenge modern AI systems based on unavailable knowledge. That is not CHALMERS ET AL.'S point. Their main argument is a problem of direction of experiments and course of action. They claimed, at the time, that enough attention had not been paid in AI research to the conceptual differentiation between "low-level perception" and "high-level perception", let alone the bridging of the gap.[35]:

> "On one side of the barrier, some models in low-level perception have been capable of building primitive representations of the

33. See HILL ET AL. (2019: 2).
34. See CHALMERS ET AL. (1992: 186).
35. Low-level perception poses so many problems that for now, the modeling of full-fledged high-level perception of the real world is a distant goal. See CHALMERS ET AL., (1992: 202).

> environment, but these are not yet sufficiently complex to be called 'meaningful'. On the other side of the barrier, much research in high-level cognitive modeling has started with representations at the conceptual level, such as propositions in predicate logic or nodes in a semantic network, where *any meaning that is present is already built in*. There has been very little work that bridges the gap between the two (...).

(ii) Sufficient similarity

The shortcomings of "relevant similarity" in AI need not deter us from assessing the equally difficult concept of "sufficient similarity". "Sufficient similarity" is a tricky concept. Many have written about it but ultimately it inevitably boils down to a matter of decision-making. And it does make sense as it was mentioned above that generalization is said to be the key distinction between decision-making and categorization.[36].

Consider these two quotations, from HAGE (2005: 408) and RODRIGUEZ (2001: 140):

> "One can *decide* that the facts b and e are sufficiently similar and consequently that the two cases also count as sufficiently similar to conclude that NC [new case] provides precisely as much support for C as OC [old case]."
>
> (Hage)

> "The final difficulty that must be addressed is that of determining a cut-off as to when two cases are similar enough. Is an 80 per cent match enough of a similarity? Why? How many cases constitute a solution? *All above 80 per cent similarity*, or only the best-matching one? Currently, these parameters must be set manually by the knowledge engineer and domain expert. Obviously, the accuracy of their estimate may affect the overall performance of the system".
>
> (Rodriguez)

As stated, "these parameters [of sufficient similarity] must be set manually by the knowledge engineer and domain expert". The threshold upon which sufficient similarity is obtained and a category is formed is a matter of human manipulation. But this human decision-making ought to have some kind of criterion. However conventional the criterion for categorizing is, one can wonder whether (and when) ROSCH's two psychological principles of categorization may be programmed into AI Systems: (i) the goal of mirroring the underlying structure in the perceived world (or legal system) combined

36. See SEGER, PETERSON (2013: 1189).

with (ii) cognitive economy would evidently impact and finetune the level of abstraction of categories that *ought to be* formed by the AI System.

On the one hand, the AI System would aim at better cohere with the underlying dispositions of things or rules (*e.g.,* that wings co-occur more often with feathers than with fur or that expressing oneself co-occurs more often with a permission than with a prohibition). On the other, the AI System would be governed by cognitive economy as a way to "to provide maximum information with the least cognitive effort" (Rosch 1978), thus avoiding optimal solutions that are extremely computationally expensive. The threshold of "sufficient similarity" would, therefore, not be manually set at 80% or 90% but precisely at the optimal point of balancing between maximum information with the least cognitive effort.[37].

The way to go about drawing up these criteria is the subject for a different paper.

## 6. Conclusions

The purpose of this paper was to outline some relevant theoretical discussions surrounding similarity and legal analogy, both in general and as particularly applied to AI. I began by outlining the main reasons why it is said that current AI systems are far from deploying the capability of forming humanlike abstractions or analogies: the fact that AI systems do not experience stuff, the problem of KEB, the difficulties in encoding different goals for analogical reasoning into AI systems and the failure to accurately transfer memory data to cases and situations outside their training regimes.

Upon laying down some fundamental terms and definitions, I focused on the lack of consensus regarding the theoretical background theories of both general analogy and AI analogy. I described the inferential model of analogy while adding some amendments to prior models and underlying some fundamental differences, notably between testable and non-testable scenarios. With this I aimed at singling out the specific problems of normative analogies, ranging from the non-testability to the paradox of non-redundancy and the problem of justification. Subsequently, I attempted to clarify the inferential steps in analogical reasoning, of which the abduction takes the center stage.

I have subsequently moved on to expound on the feature-based model of similarity – a necessary yet insufficient condition of analogy – while proposing some adjustments suggested in the expert literature, such as alignment and correspondence in asserting similarity facts. I attempted at showing that it is not instrumentally irrational to categorize as equals objects *a* and *b*, and to leave *c* out, even when *a* and *c* are more similar than *a* and *b*. Moreover, I intended to lay down the fundamental normative and

37. "[O]ne purpose of categorization is to reduce the infinite differences among stimuli to behaviorally and cognitively usable proportions. It is to the organism's advantage not to differentiate one stimulus from others when that differentiation is irrelevant to the purposes at hand" (Rosch: 1978: 29).

psychological discussions as a general problem, in order to apply it to the specific context of AI. My main examples were Tversky's directionality and diagnosticity principles, perhaps some good insights for the current debates about "sticky" stereotypes and discriminatory treatments attributed to AI systems.

I have finished off by taking stock and addressing the two specific problems of "relevant similarity" and "sufficient similarity" in AI. Regarding the former, I recalled Chalmers et al.'s distinction between low-level perception and high-level perception as well as the criticism made, dating from 1992, to research in AI that worked purely on knowledge representation for the inputs based on a notation which was already presented in the style of predicate calculus. This plea not to disregard the perceptual processes still proves important to current research on AI. Lastly, on the difficult topic of "sufficient similarity", I stressed, once again, the decision-making aspect of categorization while suggesting, *pace* ROSCH, that it ought to be governed by some principles. I have claimed that programming ROSCH's two psychological principles of categorization (as an example) into AI Systems would helpfully provide for flexible criteria to finetune the level of abstraction of categories formed in analogy-making.

**References**

Adler JE. Asymmetrical Analogical Arguments. Argumentation. 2007 21: 83–92.

Araszkiewicz M. Analogy, Similarity and Factors. In Proceedings of the 13th International Conference on Artificial Intelligence and Law, New York. 2011; ACM: 100–105.

Canale D, Tuzet G. The A Simili Argument: An Inferentialist Setting, Ratio Juris. 2009, 22: 499–509.

Chalmers D, French R, Hofstadter D. High-level perception, representation, and analogy: A critique of artificial intelligence methodology. Journal of Experimental and Theoretical Artificial Intelligence. 1992, 4 (3): 185–211.

Comanducci P. Igualdad in Democracia, Derechos e Interpretación Jurídica – Ensayos de Teoria Analítica del Derecho, P. Comanducci (Auth.), Ara Editores, Lima, 2010: 29–44.

Cumpa J. The Neutralist Analysis of Similarity in Philosophia - Philosophical Quarterly of Israel. 2021, 49 (1): 37–47.

Davies T. Determination, Uniformity, and Relevance: Normative Criteria for Generalization and Reasoning by Analogy. D.H. Helman (Ed.), Analogical Reasoning – Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy, Springer, Dordrecht, 1988: 227–251.

De Corte E, Verschaffel L, Van Dooren W. Heuristics and problem solving, Encyclopedia of the Sciences of Learning, Springer, 2012: 1422–1425.

Echave TD, Urquijo ME, Guibourg RA. Lógica, Proposición y Norma, 7th reprint, Astrea, Buenos Aires, 2008.

Endler J. Natural Selection in the Wild. Princeton University Press: Princeton, NJ, 1986.

Gianformaggio L. "Like" – "Equal" – "Similar": Are They to be Treated Alike? in Normative Systems in Legal and Moral Theory, Festschrift for Carlos E. Alchourrón and Eugenio Bulygin, E. Garzón Valdés/W. Krawietz/G. H. Von Wright/R. Zimmerling (Eds.), Duncker & Humblot, Berlin, 1997: 251–278.

Goldstone RL, Son JY. Similarity. KJ Holyoak, RG Morrison (Eds.), The Cambridge Handbook of Thinking and Reasoning. Cambridge: Cambridge University Press, 2005: 13–37.

Hage J. The Logic of Analogy in the Law. Argumentation. 2005, 19: 401–415.

Hart HLA. The Concept of Law. 2nd ed. Oxford: Oxford University Press, 1994.

Hill F, Santoro A, Barrett DGT, Morcos A, Lillicrap T. Learning to make analogies by contrasting abstract relational structure. ICLR, 2019: 1-18.

Holyoak KJ. Analogy. In The Oxford Handbook of Thinking and Reasoning, Oxford University Press, New York, 2012: 117–142.

Hummel JE, Holyoak KJ. Distributed representations of structure: a Theory of Analogical Access and Mapping. Psychological Review, 1997: 104 (3): 427-466.

Ifenthaler D. Measures of Similarity. Encyclopaedia of the Sciences of Learning. Boston, Springer, 2012: 2147–2150.

Komatsu LK. Recent views of conceptual structure, Psychological Bulletin. 1992, 112: 500–526.

Liao TW, Zhang Z, Mount CR, Similarity measures for retrieval in case-based reasoning systems, Applied Artificial Intelligence, 1998, 12: 267–288.

Lopes, PM. Relevance and Equality: an Analytical Account. The Journal of Contemporary Legal Issues (University of San Diego), 2022: 331–382.

Mandal A, Ghosh K, Ghosh S, Mandal S, Unsupervised approaches for measuring textual similarity between legal court case reports, Artificial Intelligence and Law, 2021, 29 (1): 1–35.

Maranhão J. Positivismo Jurídico Lógico-Inclusivo, Madrid: Marcial Pons, 2012.

Marinucci L, Mazzuca C, Gangemi A. Exposing implicit biases and stereotypes in human and artificial intelligence: state of the art and challenges with a focus on gender, AI & Society, 2022: 1–15.

Mitchell M. Abstraction and Analogy-making in Artificial Intelligence – Ann., NY Acad. Sci. 2021: XXXX (21): 79–101.

Murphy GL, Medin DL. The role of theories in conceptual coherence. Psychological Review. 1985, 92: 289–316.

Peczenik A. On Law and Reason, 2.ª ed., Springer: Dordrecht, 1989.

Ratti GB. El Gobierno de las Normas, Madrid, Marcial Pons, 2013.

Raz J. Practical Reason and Norms. 3rd ed. Oxford: Oxford University Press, 1999.

Rissland E. Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning, The Yale Law Journal. 1990, Vol. 99, No. 8 (Jun.): 1957–1990.

Rissland E. Using CBR to Drive IR, IJCAI'95: Proceedings of the 14th international joint conference on Artificial intelligence. 1995, Volume 1, August: 400–407.

Rissland E. AI and Similarity, Intelligent Systems, IEEE. 2006, 21(3): 39–49.

Rodríguez A. Issues in case-based reasoning. Hahn U / Ramscar M (eds), Similarity and Categorization. Oxford: Oxford University Press, 2001: 131–150.

Rosch E. Principles of Categorization. Rosch E / Loyd BB (Eds.), Cognition and Categorization, New York: John Wiles & Sons, 1978.

Russell B. On the relation of universals and particulars. Proceedings of the Aristotelian Society. 1911, 12: 1–24.

Schauer F. Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life. New York: Oxford University Press, 1991.

Schvaneveldt R. Abductive reasoning. Encyclopedia of the Sciences of Learning, Boston: Springer, 2012: 12–15.

Searle J. A Classification of Illocutionary Acts in Language and Society. 1976, 5 (1): 1–23.

Seger CA, Peterson EJ. Categorization = decision making + generalization In Neurosci Biobehav Rev. 2013, 37(7): 1187–1200.

Strawson PF. Subject and Predicate in Logic and Grammar, Methuen, London, 1974.

Thagard P, Holyoak KJ, Nelson G, Gochfeld D. Analog Retrieval by Constraint Satisfaction In Artificial Intelligence, 46 (3), 1990: 259–310.

Tversky A. Features of similarity. Psychology Review. 1977, 84: 327–352.

Tversky A, Gati I. Studies of Similarity. E. Rosch / B. B. Loyd (Eds.), Cognition and Categorization, New York: John Wiles & Sons, 1982.

Yovel J. Two Conceptions of Relevance. In Cybernetics and Systems. 2003, 34:4-5: 283–315.

[i] This paper was prepared within the framework of the Horizon Twinning project "Advancing cooperation on The Foundations of Law – ALF" (project no. 101079177). The project is financed by the European Union.