Forthcoming in B. Ball, A. Helliwell, A. Rossi (eds.), *Wittgenstein and AI* (vol. I), Anthem Press.

Modelling Analogical Reasoning: One-Size-Fits-All?

Ioannis Votsis

1. Introduction

A key type of reasoning in everyday life and science is reasoning by analogy. Roughly speaking, such reasoning involves the transposition of solutions that work well in one domain to another, on the basis of pre-existing shared properties between the two domains. If we are to automate scientific reasoning with artificial intelligence (AI), then we need adequate models of analogical reasoning that clearly specify the conditions under which good analogical inferences can be made and bad ones avoided. Two general approaches to such modelling exist: universal and local. In this chapter, we assess the merits and demerits of both approaches. We concede that there are substantial obstacles standing in the way of the universal model view, but that these may be mitigated to some extent by supplementing existing models with additional criteria. One such criterion is defended, particularly against a challenge due to Wittgenstein. We argue that this challenge can be met and thus that there is hope for a one-size-fits-all model in the study of analogical reasoning.

The structure of the chapter is as follows. Section 2 provides an overview of the main philosophical models of analogical reasoning, identifying some of their strengths and weaknesses. Section 3 briefly looks at one model of analogical reasoning that originates in the symbolic AI tradition, and offers some very general remarks about the prospects of modelling analogical reasoning with neural AI. Section 4 sets out the key issue of concern for this chapter, namely whether a universal model of analogical reasoning can be constructed. Section 5 considers one promising route towards a universal model via the criterion that the concepts involved are relevantly uniform. Section 6 presents a challenge to this route that can be found in Wittgenstein's family resemblance metaphor, whose ultimate target is the rejection of concept uniformity. An attempt is made to meet this challenge by arguing that some concepts in natural science are uniform, or at least more uniform than others, but also that scientific inquiry strives towards, and manages to increase, uniformity. Section 7 highlights the testability of the proposed criterion as well as the relative ease with which it can be computationally implemented, raising the overall prospects of automating the process of scientific discovery. The essay concludes with Section 8, which contains a summary of the main points but also a parting attempt to answer the question why analogical reasoning works at all.

2. Philosophical Models of Analogical Reasoning

In this section, we explore some of the main models of analogical reasoning, particularly as they are applied to the sciences. These models are taken from the philosophical literature. Let us begin the discussion with some useful terminology.

Analogical reasoning is reasoning that exploits analogies. Performing such reasoning first requires some known or accepted similarities between the properties and/or relations – henceforth, simply 'properties' – of two domains, which we may call 'source' S and 'target' T. These similarities ground the analogy, which is then employed to infer an additional similarity between S and T. The additional similarity concerns a property known or accepted to hold in S, but heretofore not known or accepted in T. Darwin posited his theory of evolution by natural selection by, among other things, drawing inspiration from two sources: (1) artificial selection and (2) Malthus' principle of population growth. Artificial selection involves the breeding of animals or plants to suppress or accentuate certain traits. During Darwin's time, for example, pigeons and other birds were bred for exhibition purposes by making their beaks smaller and their chests bigger. This was accomplished by successively mating birds that exhibited the desired traits, which were then inherited in the next generation, leading to a slow but steady tendency towards those traits. Malthus's principle of population growth was an attempt to model what happens to the size of a human population when the availability of resources varies. In times of plenty, and other things being equal, such populations grow. At some point, however, the growth outpaces the availability of resources, leading to competition and a struggle for existence, including war and population reductions. Using analogical reasoning, Darwin argued that nature places similar selection and resource pressures on animals and plants. These pressures suppress or accentuate traits depending on whether or not possessing them provides advantages or drawbacks in the struggle for survival and reproduction. In his own words:

> [w]hy may I not invent hypothesis of natural selection (which from the analogy of domestic productions [i.e. artificial selection], & from what we know of the struggle of existence [i.e. Malthus' principle] & of the variability of organic beings, is in some very slight degree, in itself probable) & try whether this hypothesis of natural selection does not explain (as I think it does) a large number of facts. (Letter to J.S Henslow, quoted in Darwin [1860] 1967: 204).

Analogical reasoning is typically construed in argument form. The premises assert known/accepted similarities between the properties or relations of S and T as well as some additional property that S is known/accepted to possesses. That additional property is then inferred in the conclusion to be true of T. As an example, we may turn to the abovementioned analogy, whose corresponding argument may be formulated as follows (where AS is Artificial Selection, NS is Natural Selection, MP is Malthus' Principle):

1. AS is similar to NS in that selective reproduction affects which traits are inherited.
2. MP is similar to NS in that there is competition for resources.
3. AS selection pressures affect survival and reproduction rates.
4. MP resource pressures affect survival and reproduction rates.
5. Therefore, NS selection and resource pressures affect survival and reproduction rates.

Stated thus, the argument is best characterized as either inductive or abductive in form. Still, it can be turned into a deductive argument by finding and stating the missing premises. Two questions arise here. One is descriptive: What form of arguments are involved in analogical reasoning as it is practiced? The other is prescriptive: What form of arguments should be involved in such reasoning? We shall not wade into this debate. For simplicity, and unless otherwise noted, the analogical arguments presented hereafter are cast in broadly inductivist terms.

Another important issue that comes up in this literature concerns the role of analogical arguments. Sometimes such arguments are utilized to provide support for a conclusion, thereby playing a justificatory role. Indeed, as Bartha notes: "In fields such as archaeology, where we lack direct means of testing, they may provide the strongest form of support available" (2010: 2). More frequently, however, they are utilized to provide some initial plausibility, thereby playing a discovery or heuristic role. The former role is typically thought to be more demanding than the latter. As a result, models of analogical reasoning that are geared towards justification are also employed for discovery but not vice versa. In what follows, we are focusing more on discovery rather than full-blown justification. This stance is not meant to prejudge our attitude towards the debate over the proper role for analogical reasoning. To make this clear, whenever possible, we will employ the neutral term 'admissible' to denote the various roles such arguments may play. That is to say, admissibility may be enunciated in different ways, e.g. example, initial plausibility, probability, and so on.

Let us explore some philosophical models of analogical reasoning. The first such model is simplistic but useful in that it conveys some basic ingredients that go into modelling analogical reasoning:

The Simple Schema (TSS)

An analogical inference from S to T (in relation to feature Q) is admissible if and only if:

"(1) S is similar to T in certain (known) respects.
(2) S has some further feature Q.
(3) Therefore, T has the feature Q, or some feature Q* similar to Q." (Bartha 2010: 13).

This is a no-frills version of analogical reasoning. The key question is whether Q, the additional feature of S asserted in premise 2, is also a feature of T. Given the analogy between S and T established in premise 1, we are allowed to conclude that it is also a feature of T. Note that, strictly speaking, the inferred feature need not be identical but may just be merely similar to Q – hence the reference to Q*.

TSS is clearly too liberal. Almost any S–T pairing can be deemed to be analogous and therefore ripe for an extended analogy to a corresponding feature Q. That's because the only requirement TSS brings to the table is the existence of some similarities between S and T, and these are rather easy to come by. Here's a made-up example that illustrates how this liberality can lead to absurd results:

1. Diamonds and chalk are made up of carbon-based molecules.
2. Diamonds have a hardness rating of 10 on the Mohs scale.
3. Therefore, chalk has a hardness rating of 10 on the Mohs scale.

This is obviously a bad inference to draw as chalk is a very soft rock. It is made up of calcite, which actually has a rating of 3 on the Mohs scale, where 10 is the hardest and 1 the softest.

Another philosophical model, one that seeks to plug the holes left behind by TSS, is Hesse's (1966) causal model. The model is meant to establish justification for the inferred similarity by imposing three restrictions on analogical arguments. First, the known/accepted similarities between S and T that set up the basic analogy must be observational (the observational condition). Second, the properties in S that form part of the basic analogy must be causally connected to the additional property in S that forms part of the extended analogy (the causal condition). Third, there must not be essential or causal properties in S that are known to be dissimilar with essential or causal properties in T (the no essential dissimilarity condition). This model is typically presented in tabular form:

Hesse's Causal Model (HCM)

Similarities \ Domains	S	T
Known Observational Similarity	Property Q_1	Property Q_1 (or Property Q_1*)
	…	…
	Property Q_n	Property Q_n (or Property Q_n*)
Inferred Observational Similarity	Property Q_{n+1}	Property Q_{n+1} (or Property Q_{n+1}*)

The known similarities, which establish the basic analogy, are conceived of as horizontal relations (denoted by the double-headed arrow) between the observational properties Q_1, …, Q_n of S and those of T. The causal relation between Q_1, …, Q_n and Q_{n+1} is conceived of as a vertical relation (denoted by the single-headed arrow). Finally, the inferred similarity is property Q_{n+1} (or Q_{n+1}*), which is located under T at the very bottom on the right. The last row establishes the extended analogy between S and T. Applying this model to the diamond–chalk case, we can block the inference that chalk has a hardness rating of 10, because the property of being made up of carbon-based molecules on its own does not cause an object to have that level of hardness.

Despite generally acknowledged as a step in the right direction, HCM is seen as ultimately inadequate. One common accusation is that it is too strict because it requires a causal connection between Q_1, …, Q_n and Q_{n+1} where (presumably) sometimes a mere correlation can serve just as well. As an example, Bartha (2010) gives Benjamin Franklin's inference that (metal) rods attract lightning in the wild just like they attract electrical fluid in the lab, which Franklin bases on several existing similarities between electrical fluid and lightning: "1. Giving light. 2. Colour of the light. 3. Crooked direction… 10. Melting metals. 11. Firing inflammable substances. 12. Sulphureous smell." (Franklin 1941: 334). This was a good analogical argument, Bartha argues, even though "[t]here was no known causal connection between the twelve 'particulars' [the known similarities] and the thirteenth property [the inferred similarity], but there was a strong correlation" (2010: 44).

An altogether different philosophical approach puts probability front and centre, seeking to quantify the goodness of analogical arguments. In a nutshell, known/accepted similarities increase the overall likeness between S and T, thereby lending more support to the inferred similarity. The idea goes back to John Stuart Mill ([1843] 1973), who argued that 'There can be no doubt that every such resemblance which can be pointed out between B [read: T] and A [read: S], affords some degree of probability, beyond what would otherwise exist, in favour of the conclusion drawn from it' (p. 556). The properties that those resemblances are about, he explained, 'must not be properties known to be unconnected with it' (p. 555). That is, they must not be properties that are not known to be irrelevant
for the extended analogy. We can give this model a modern formulation, interpreting probabilities as rational degrees of belief, a.k.a. 'credences', as follows:

Mill's Probability Model (MPM)

The admissibility of an analogical inference from S to T (in relation to property Q_k) increases if and only if: (a) $S(Q_k)$ and (b) for any property Q_i that is distinct from Q_k, $P(T(Q_k) \mid (S(Q_i) \approx_s T(Q_i)) \,\&\, B) > P(T(Q_k) \mid B)$.

where:

$P(\cdot|\cdot)$ stands for a conditional probability function
$\Phi(Q_i)$ stands for domain Φ possessing property Q_i
$\Phi(Q_i) \approx_s \Psi(Q_i)$ stands for domains Φ and Ψ being similar with respect to property Q_i, which is not recognised as irrelevant.
B stands for background knowledge

Applying this model to the electrical fluid-lighting case enables us to draw the inference that (metal) rods attract lightning in the wild just as they attract electrical fluid in the lab, provided the similarities Franklin cites do indeed increase our credence in the inferred similarity. Questions can of course be posed about the credence-inducing credentials of those similarities, but we will not pursue them here.

MPM may overcome the restrictiveness that presumably afflicts HCM, but it does so by being more generous in its attribution of goodness to analogical arguments. As such, it opens itself up to accusations of being too liberal. Similarities that are not known to, but may actually, be irrelevant are allowed to increase credence in T possessing that additional property. As an example, take the *Ligularia fischeri* (S) and the *Caltha palustris* (T) plants. These plants look very much alike and are often mistaken for one another. Suppose that for a given individual X, S is known to be edible, but they do not know if T is edible. Suppose, moreover, that X does not know whether the properties involved in the phenotypical similarities are actually irrelevant for the edibility of the plants. Following MPM, X may then tragically draw the inference that T is edible, even though it is not only inedible but poisonous. Alas, this is not just a fabricated case. Every year several people in Korea get poisoned this way.

The final philosophical model to consider here can be found in Bartha (2010: Ch. 4). His model is designed to play only a heuristic role. Moreover, the conditions it imposes are meant to be sufficient, but not necessary, for initial plausibility:

Bartha's Articulation Model (BAM)

"An analogical argument meets the requirements for prima facie [i.e. initial] plausibility if:

1. *Overlap*. $\varphi^{+} \cap P \neq \emptyset$ (where $\emptyset$ is the empty set).
2. *No-critical-difference*. $\varphi^{C} \cap N = \emptyset$." (Bartha 2010: 101).

The symbols are articulated as follows: P denotes all the properties in the positive analogy (i.e. the known similarities between S and T), N denotes all the properties in the negative analogy (i.e. the known dissimilarities between S and T), φ^{+} denotes the properties in S that causally or merely correlationally contribute to the presence of the additional property in S, and φ^{C} denotes the properties in S that are critical factors, i.e. "those elements of the prior association [i.e. the known similarities] represented as playing an *essential* part *in the circumstances*" (100) [original emphasis]. What Bartha seems to mean by the first condition is that at least one property in S used in the positive analogy between S and T must be causally or merely correlationally connected to the additional property in S. This condition is clearly made in the image of HCM's vertical relation, but a bit looser to allow for correlational, not just causal, connections. What he seems to mean by the second condition is that no property in S that plays a critical role in that connection should be part of the negative analogy between S and T. BAM is further developed in the same chapter to deal with arguments that employ multiple analogies. Moreover, in a subsequent chapter (Ch. 8), Bartha offers an adaptation of his model in probabilistic terms, which he calls 'non-negligible prior probability'. As the details of this adaptation are quite involved, we refrain from its exposition here.

The model is not without its critics. One objection concerns the sanctioning of both causal and correlational connections between the base analogy properties and the property in the extended analogy. Although BAM's inclusion of correlational cases appears to be a step forward when compared to HCM's outright prohibition, it is still unclear why we should admit all correlational cases when making such inferences. If only some correlational cases pass muster, we need criteria that distinguish the good ones from the bad ones. Put otherwise, BAM appears to be too liberal in its conception of analogical reasoning.

3. AI Models of Analogical Reasoning

Let us now discuss, albeit briefly, AI models. These are often accompanied by computational implementations and come in roughly two flavours: symbolic and neural.[1] The symbolic models include the structure-mapping engine (SME) (Falkenhainer, Forbus and Gentner 1986) and the active-symbol architecture (Hofstadter 1995). The neural models include the wild relation network (Barrett et al. 2018), and the emergent symbol binding network (Webb, Sinha and Cohen 2020). In what follows, and due to space limitations, we consider only one symbolic model, the SME, but also make some general remarks about what can be expected from neural models.

The SME is a computational model that implements Gentner's (1983) structure-mapping theory of analogical processing in cognition. The model restricts similarities to structures, i.e. the relations between elements and even the relations between relations, of the domains in question. In more detail, neither the elements nor the monadic properties of T need to resemble the elements or the monadic properties of S, but some relations in T must be similar to relations in S. Although not explicitly expressed in terms of analogical inference admissibility, we may provide such a formulation for ease of comparison with the foregoing models:

[1] Mitchell (2021) provides a useful and up-to-date overview of AI models and divides them into three types: symbolic, deep learning and probabilistic program induction.

Structure-Mapping Engine (SME)

An analogical inference from S to T with respect to relation R_k is admissible if (and only if?): (a) R_k is a relation in S, (b) there is a systematic mapping between some set of relations R^S in S and some set of relations R^T in T, where $R_k \notin R^S, R^T$ and (c) in case there is more than one such mapping, the one with the highest systematicity is prioritized as a basis for the inference.

Note that it's not clear whether the conditions in this model are envisioned to be merely sufficient, or also necessary. Note, moreover, that the critical concept in those conditions is systematicity. This, in effect, means that the mapping between relations must include higher-order relations, that is, relations between relations, because these (presumably) indicate that the knowledge in a given domain is connected.

It's worth considering a few details about the implementation, which consists of three steps. The first step, roughly, involves the search for all possible individual relation and object pairings between S and T. The relations are described in a logical or quasi-logical language that contains constants and predicates. These predicates are nested in a tree-like structure to form expressions. To understand what's going on, let us adapt an example from Falkenhainer, Forbus and Gentner (1986). Suppose that the domains in question are the solar system S and the Rutherford atom T. S is described in terms of the following nested two-place predicates: (i) Causes(And, Revolves_Around()), (ii) And(Attracts(), Greater_Mass()), (iii) Greater_Mass(Sun, planet), (iv) Attracts(Sun, planet), and (v) Revolves_Around(planet, Sun). T is described in terms of the following two-place predicates: (vi) Greater_Mass'(nucleus, electron) and (vii) Attracts'(nucleus,electron). Two object pairings can be established in this case: nucleus – Sun and electron – planet. These pairings are suggested by the mass inequality relational pairing, expressed by the similarly named predicates Greater_Mass() and Greater_Mass'(), and the attraction pairing, expressed by the similarly named predicates Attracts() and Attracts'(). The second step, roughly, involves the construction of all possible global mappings. These are mappings that merge individual pairings into a coherent whole. Finally, the third step, involves an evaluation of global mappings to select the one that scores highest on systematicity. In the case at issue, assuming no other pairings can be produced, SME will suggest that the other predicates in S also apply in T, e.g. Revolves_Around(electron, nucleus). More crucially, it will suggest that "the [mass] inequality, together with the mutual attraction of the nucleus and the electron, causes the electron to revolve around the nucleus" (275). This is, roughly, what happened in the history of atomic physics, when the Rutherford-Bohr model of the atom was proposed. According to this model, which has since been superseded, electrons are kept in orbit around the nucleus of an atom with an electrostatic force, just like planets are kept in orbit around the Sun with a gravitational force.

One objection to the SME model is that a lot hinges on the choice of predicates. Not only is that choice difficult to make, but there are also questions about the effects it has on the drawing of analogical inferences. The latter issue generalizes into a problem that afflicts various areas of philosophy and beyond. That judgements concerning such issues as the ordering of theories on the basis of verisimilitude or simplicity may be affected by linguistic choices is a decades-old problem (Miller 2017). This problem is further compounded by the fact that SME does not provide any guidance on how to choose or indeed construct predicates, but assumes that these have already been supplied. Additional problems with SME are discussed in Bartha (2010: Ch. 3), who complains that systematicity is neither necessary nor sufficient for good analogical inferences.

It is now time to turn to some general remarks about neural net models of analogical reasoning. Neural net approaches to AI have been on the ascendancy in recent years. A big part of the reason why is the fact that deep neural nets, unlike symbolic systems, are more readily able to handle data that are not fully structured. As Kautz (2022) stresses, deep neural nets obviate "the need for manually engineered features [machine learning speak for variables]" (112). This opens opportunities that were once unavailable. Still, one major obstacle in the way of neural models of analogical reasoning is that neural nets are not particularly good at reasoning. Their hidden layers of nodes, where the computation occurs, are notoriously difficult to interpret. In fact, not only do those computations fail to resemble human reasoning, but they also often require vast amounts of data to draw simple inferences.

Despite these obstacles, recent attempts at making headway on the problem of computationally modelling analogical reasoning have leaned heavily on neural nets. On this approach, analogy making is something that needs to be learned. This involves feeding neural nets with relevant training data, e.g. correct and incorrect analogies. To address the problem that the data needs to be orders of magnitude higher than the available correct and incorrect analogies, AI theorists and practitioners have resorted to the automatic generation of data. In some areas, this process is easier to carry out than in others. Non-verbal analogical reasoning tasks involving shapes whose attributes (e.g. colour and shape) vary, known as Raven's progressive matrices, are now routinely explored with neural nets, precisely because the production of vast amounts of training and testing data can be automated.

Although neural nets trained on such sets are gradually getting better at drawing the right analogies, deep disagreements have emerged (Barrett et al. 2018; Hu et al. 2021; Zhang et al. 2019) over whether the automatically produced data are diverse enough to give rise to sufficiently stringent tests of the resulting models. Moreover, and as already indicated above, not all reasoning tasks are easily amenable to the artificial generation of data. Given the generally higher complexity of analogical inferences in science (vs. in non-verbal tasks), one would expect that artificial data would be harder to synthesize *ab initio*. As such, analogical reasoning in science presents a significant challenge for those advocating neural net approaches. There is also a more general reason to doubt the amount of mileage we can get out of neural nets. If we had a method to produce diverse data that provide sufficiently stringent tests for our models, then we wouldn't really need (to test) those models because the method itself would presumably generate the desired model or something like it.

It is not our intention here to say that neural net models face insurmountable problems as regards the modelling of analogical reasoning in science. Rather, we just wanted to highlight some genuine difficulties. To end on a more positive note, we would like to point out that neuro-symbolic approaches to AI are increasingly being adopted to solve problems that require the complementary strengths of neural nets and symbolic methods. Kautz (2022), for example, reports that 'the next big scientific advance in AI' will involve such hybrid systems. We expect a similar tendency in the area of analogical reasoning in science, raising the prospects of a fully automated approach to scientific discovery.

We conclude this section with some big picture remarks on the analogical reasoning models, both philosophical and AI, on offer. Such models are prescriptive in that they do not merely, or primarily, concern themselves with the actual practice of discriminating between good and bad analogical reasoning. Some of them are explicitly qualitative (TSS, HCM, non-probabilistic BAM), while others are quantitative (MPM, probabilistic BAM, SME). Some are merely heuristically-oriented (BAM, SME), hoping to establish initial plausibility claims about inferred similarities. Others play both a heuristic and a justificatory role (TSS, HCM, MPM), hoping to also establish support towards those claims. All of them encounter counterexamples, which allege either that they are too liberal (TSS, MPM, SME, BAM) or too stringent (HCM).[2] Despite these drawbacks, nothing of what we said here prevents their amelioration through a modification (addition, deletion, or both) of their stated conditions. The section following the next offers one such additional criterion that can be bolted on to any of the above models. Before we get to that section, however, a general challenge awaits.

[2] Although we have not discussed this point earlier, it is worth noting that some models have been accused of being too strict in some respects and too liberal in others.

4. Norton's Material Challenge

An easy way out of having to deal with counterexamples is to deny that any such models, even when they are modified, cover all (and only) instances of good analogical reasoning. This is tantamount to denying the existence of a universal model of analogical reasoning, and it is a route that several scholars have taken. Some (Currie 2013; Reiss 2015; Toulmin 1958) have suggested that different fields of research require different models. Others, like Norton (2003, 2011, 2021), have suggested a more fine-grained approach, claiming that different research questions, even when these emanate from the same field, require different models. In this section, we consider a general challenge against those who wish to analyse analogical modelling in terms of a one-size-fits-all, that is, a universal model.

The challenge is due to Norton and is based on the claim that inductive inferences are only good in so far as they are licensed by local facts pertaining to the situation. His argument is based on an ingeniously simple example. Compare the following two inductive inferences:

	Inference 1	**Inference 2**
Premise:	Some samples of bismuth melt at 271 °C.	Some samples of wax melt at 91°C.
Conclusion:	All samples of bismuth melt at 271 °C.	All samples of wax melt at 91°C.

Although the two inferences are structurally identical, only the first one is good. That's because all samples of bismuth do melt at 271 °C, but not all samples of wax melt at 91 °C. On Norton's view, the goodness of an inductive inference is "grounded in matters of fact that hold only in particular domains" (2003: 647). Thus, there are facts about bismuth, but not about wax, that make it the case that all of its samples behave in the same way.

Unsurprisingly, Norton extends this attitude to analogical reasoning, which he considers to be a species of inductive reasoning. In doing so, he repudiates the universal model of analogical reasoning:

> If analogical reasoning is required to conform only to a simple formal schema, the restriction is too permissive. Inferences are authorized that clearly should not pass muster.… The natural response has been to develop more elaborate formal templates that are able to discriminate more finely by capturing more details of various test cases… elaborations cannot escape the inevitable difficulty. Their embellished schema [is] never quite embellished enough. There is always some part of the analysis that must be handled… without guidance from strict formal rules. (2021: 119-120).

In other words, he takes the whole project of attempting to construct models like the above as doomed from the outset. Such models face the impossible (to him) task of trying to catch all the counterexamples with the introduction of more and more qualifications. Expressed a different way, modellers may be facing the impossible task of trying to find an elusive balance between just the right amount of permissiveness and just the right amount of restrictiveness.

The compelling force of the bismuth-wax example notwithstanding, it is worth asking whether the process of adjusting models of analogical reasoning has no end in sight. Is the trade-off between maximizing correct analogies and minimizing incorrect ones necessarily unavoidable? To definitively answer questions like these in the affirmative is no trivial matter, and would require an impossibility proof. No such proof has been given. At best, what Norton presents us with is an inductive case (whose own licencing fact is not entirely justified) for pessimism vis-à-vis universal models of analogical reasoning. In the sections that follow, we want to give some hope to the optimists by motivating an additional criterion of admissibility for analogical reasoning. This criterion may be combined with any of the aforementioned models.

5. Relevant Conceptual Uniformity

In this section, we claim that relevant conceptual uniformity is a fruitful, potentially even necessary, additional criterion for the determination of good from bad analogical arguments. To set up this criterion, we first need to say something about relevance, uniformity, and the relation between them.

To say that a concept is unqualifiedly uniform is to say that the things/tokens it represents are homogeneous with respect to some of their natural properties. The uniformity required is not total homogeneity of the tokens, that is, sameness of all their properties, for the simple reason that that would result in only one token per concept. Moreover, conceptual uniformity is not to be used in isolation from, but rather in tandem with, relevance constraints. It should be clear, from the diamond–chalk example, that to avoid drawing unsuitable analogical inferences, such constraints must be in place. These, as we have seen above, may take the form of causal or correlational conditions relating the properties (and concepts) in the base analogy to the property (and concept) in the extended analogy. A set of properties (and its corresponding set of concepts) is unqualifiedly relevant for another set of properties (and its corresponding set of concepts) if and only if the former set of properties completely and non-redundantly determines the presence (and any values) of the latter. As such, the former can be employed to genuinely explicate lawlike behaviour relating to the latter. For example, certain genetic properties of *Ligularia fischeri* as well those of humans are presumably causally relevant in genuinely explicating that plant's edibility by humans. The claim we would like to put forward here is that if the former properties and concepts are unqualifiedly relevant for the latter properties and concepts, then those concepts must be unqualifiedly uniform. Another way of expressing this relationship is that unqualified conceptual uniformity is a necessary, but not sufficient, condition for unqualified conceptual relevance.

Focusing only on unqualified conceptual uniformity ignores the fact that various concepts are gainfully employed in scientific reasoning but are not entirely uniform – the concept species is a well-known example. Indeed, as we will see in the next section, scientific concepts may start life as fairly dis-uniform, and gradually build towards increased uniformity. That means it is worth considering the extent to which a concept is uniform. Similarly, focusing only on unqualified property/conceptual relevance ignores what we have already implicitly conceded, namely that it is worth considering the degree to which a property/concept is relevant to another property/concept. This is obvious in cases where the base analogy properties/concepts may be imperfectly correlated with the extended analogy property/concept – think of the electrical fluid-lightning case. Taking these observations into account, we may say that the concepts featuring in the analogy between S and T must be restricted to those that are relevantly uniform. That is, they must be concepts whose tokens exhibit a certain degree of homogeneity vis-à-vis some of their natural properties, with the properties involved in the base analogy at least partly determining the presence (and any values) of the property in the extended analogy. Other things being equal, increasing the strength of that determination relation should lead to a corresponding increase in the degree of homogeneity between the tokens. Based on these ideas, we can then define the following criterion:

> **Relevant Conceptual Uniformity Admissibility Criterion**: Other things being equal, the more relevantly uniform those concepts, the higher the admissibility of the analogical inference.

To see the usefulness of the notion and corresponding criterion of relevant conceptual uniformity, we need go no further than Norton's bismuth-wax example. Ironically, Norton's analysis of what really goes on in this example brings out the importance of this notion to the surface. In his own words: "All samples of bismuth are uniform just in the property that determines their melting point […] Wax samples lack this uniformity in the relevant property, since "wax" is the generic name for various mixtures of hydrocarbons" (2003: 650). In other words, it's no wonder that the inference from some to all tokens of bismuth is reliable, but the one from some to all tokens of wax is not. Bismuth, qua a chemical element, is highly uniform with respect to several properties, including, most relevantly, those that result in the lawlike behaviour of melting points.[3] In more detail, melting points are decided by how much energy is needed to overcome the intermolecular forces that make up the internal structure of a substance. Since the internal structure of different bismuth tokens is identical, the energy required is the same in all cases. By contrast, wax is not a relevantly uniform concept, at least not with respect to the internal structure of its tokens. As such, no inference to the melting point of all its tokens can be secured from some of them. Indeed, even the subordinate concept of paraffin wax, represents tokens whose melting points vary considerably because the corresponding intermolecular forces vary considerably (Himran, Suwono and Mansoori 1994).

The proposed relevant conceptual uniformity admissibility criterion can be integrated into any of the existing models of analogical reasoning. We only have space for two quick demonstrations here, so we restrict most of our comments to the TSS and MPM models. For expedience, we may speak directly of concepts being relevantly uniform with respect to other concepts, dropping the reference to properties being relevantly uniform. Moreover, we here treat the additional criterion as necessary, though this need not be the case. We may thus aptly modify these accounts as follows:

TSS (with relevant conceptual uniformity)

(1) S is similar to T in relation to properties $Q_1, \ldots, Q_n$, which are encoded by concepts $C_1, \ldots, C_n$.
(2) S possesses some further property Q_{n+1}, encoded by concept C_{n+1}.
(3) Concepts $C_1, \ldots, C_n$, are relevantly uniform vis-à-vis concept C_{n+1}.
(4) Therefore, T possesses property Q_{n+1}, or some property similar to Q_{n+1}.

MPM (with relevant conceptual uniformity)

The admissibility of an analogical argument from S to T in relation to property Q_k increases if and only if: (a) $S(Q_k)$ and (b) for any concept C_i (corresponding to property Q_i) that is distinct from, but relevantly uniform with respect to, concept C_k (corresponding to property Q_k), $P(T(Q_k) \mid (S(Q_i) \approx_{s*} T(Q_i)) \& B) > P(T(Q_k) \mid B)$.

The subscript s* in the modified MPM account signifies that S and T are similar to each other, without needing to specify that the similarity must not be known to be irrelevant. That's because the relevant conceptual uniformity criterion now carries the burden of determining relevance.

Before we bring this section to a close, it is worth appraising what it is we have addressed in the challenge posed by Norton. First of all, we must grant that uniformities cannot simply be stipulated. They must be discovered. So, Norton is right in asserting that facts of the matter enter the determination of analogical inference admissibility. Having said this, the relevant conceptual uniformity admissibility criterion is blind to the specific research question (or field) pursued. As such, it is not a local, but a universal, condition, or at least aspires to be one provided a version of it works in all cases. If such a version does indeed work in all cases, then, by integrating it into existing models of analogical reasoning, it brings us one step closer to universality. Less polemically, such models are perhaps not as local as Norton would have us believe.

6. A Wittgensteinian Spanner in the Works?

One major obstacle to the above approach has its origins in Wittgenstein (1953), where we are urged to move away from the view that terms or concepts possess essences, which can be captured by formal definitions given in terms of necessary and sufficient conditions. Language, on Wittgenstein's view, doesn't work like that. Taking the concept of games as an illustration, he argues that there is so much

[3] The periodic table of elements, of which bismuth is a member, contains some of the most uniform concepts found in nature, second only to the uniformity that exist across subatomic particle concepts.

variation in its tokens, e.g. ball games, board games, card games, etc., that it is pointless to try to find a definition. But that doesn't mean that different tokens of games are not more similar to each other than they are to other things. Using this claim as a basis, he then asserts that: "I can think of no better expression to characterize these similarities than 'family resemblances'… 'games' form a family." (§67). Norton's and Wittgenstein's rationales are similar as they both seem to claim that we philosophers are unhealthily preoccupied with 'generality', and, in so doing, omit practice, which is often grounded in local peculiarities – in the present case, facts about the usage of concepts and terms.

Wittgenstein's anti-definitional/anti-essentialist stance is in stark contrast to the requirement of (relevant) conceptual uniformity, for his view effectively denies the uniformity of concepts. On this view, tokens at best exhibit varying degrees of similarity to one another, and concepts are thus less than perfectly uniform. This presents a challenge to the limit case of the relevant conceptual uniformity
Admissibility criterion, that is, the case where the concepts involved are perfectly uniform, as it judges it to be unsatisfiable. In what follows, we consider what can be said to address this challenge.

Let us start by conceding that much of what goes on in language is as Wittgenstein describes it. Many of our concepts have meanings and extensions with unclear boundaries and are best treated in terms of graded membership. Just because many concepts are like this, however, doesn't mean that all concepts, including scientific ones, should be treated this way. That is to say, we must not infer that all concepts are dis-uniform or less than perfectly uniform from the (admittedly reasonable) claim that many everyday concepts do exhibit varying degrees of dis-uniformity. This would be as bad an inference by analogy, as the ones that the analogical reasoning modellers are so desperate to avoid.

More positively, we can, in fact, argue for uniformity from within, that is, by following Wittgenstein's own methodology. On this methodology, membership in a concept is decided through language use. That some concepts in science, particularly those in natural science, are uniform is evidenced by such use. Definitions are demanded, given and consistently followed in science. Some of the best-known examples concern the base concepts employed in the International System of Units (SI): metre, second, mole, ampere, kelvin, candela and kilogram. Incidentally, Wittgenstein (1953: §50) is unconvinced by these concepts and launches the following complaint about the standard metre. It is meaningless, according to him, to ask if the standard metre is a metre long, because the standard metre is a physical object and thus cannot be laid next to itself. This complaint is obviously outdated. At the time, the standard metre was defined against a physical object, namely a platinum-iridium bar. Today, all seven base units concepts are given definitions in terms of fundamental physical constants and each other. The standard metre is defined thus: "The meter is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second" (NIST).[4]

Some uniformity in nature is necessary. Without it, our world would be too much of a jumble to make any sense of and predict. In fact, even Hume, who suggests that causal relations and inductive inferences are but mere projections of the mind, assumes some uniformities. For, without some (restricted) uniformity of B following A, it would be impossible to form a habit of the mind that B follows A. If the world had no uniformity whatsoever, we wouldn't even be able to communicate with each other, as all categories, including sounds and words, would contain a random selection of tokens with unique features. In such a world, we would be unable to accomplish anything, unless we did so by chance.

[4] Indeed, the last remaining bastion of sample-centric standards was the standard kilogram, a platinum-iridium cylinder kept at the International Bureau of Weights and Measures in Sèrves, France. Even this however was replaced by a definition in 2019.

Assuming that there are natural categories out there, as our best (most successful) science seems to indicate, the process of accurately representing them with concepts cannot but be gradual. That's because, as with any other epistemic investigation, victories are hard-earned. Concepts must be successively refined, which, as we have already argued, involves a tendency towards greater uniformity and lawlike relevance. The fact that the SI base units, as well as various other concepts, have changed over the years, and the way they have changed, reflects this toiling process. The standard metre, for example, has changed from being equal to 1/10,000,000 of the distance that separates the equator from the North Pole to a platinum-iridium bar held at very specific temperature and pressure vacuum conditions to the contemporary vacuum traversing length of light definition. Despite these changes, and the orders of magnitude reduction in the uncertainties involved, a metre 200 years ago is still approximately the same as a metre today, at least in relation to macroscopic scales and even down to optical microscope scales.

An important consequence of refining definitions towards relevant uniformity is that doing so increases the truth(-likeness) of proposed generalizations. The generalization expressed by the sentence 'All neutrinos interact with matter to produce electrons' has some truth content in that electrons are indeed produced in neutrino interactions with matter. But it is not entirely true or even close to the truth. That's because, as it turns out, there is not just one type of neutrinos, but three: electron, muon and tau neutrinos. Only the former interact with matter to produce electrons. Thus, if we replace the concept 'neutrinos' in the above generalization with the more uniform and relevant concept 'electron neutrinos', then we end up with a generalization that is significantly more truthlike: 'All electron neutrinos interact with matter to produce electrons.' Note that the postulation of, and experimental confirmation that, neutrinos come in three different types was gradual, spanning about seven decades.

7. Testability and Computational Implementation

One beneficial aspect of the overall approach recommended in this chapter is that it is testable. Recall that the goodness of analogical reasoning, according to this approach, depends partly on the degree of relevant conceptual uniformity. As such, and other things being equal, analogical inferences are less likely to succeed when the concepts involved are less relevantly uniform. This is a prediction that drops out of the proposed approach and can be tested by running experiments with human subjects. We could, for example, assign subjects the role of discovering extended similarities on the grounds of basic similarities between different domains. If we vary the relevant conceptual uniformity levels of concepts across the subjects, we may find that higher levels are more or less helpful in making those discoveries. The puzzles could involve scenarios inspired by real scientific discoveries made with analogical reasoning, in which case it's best to select subjects with no prior knowledge of science.

Another way to test the prediction is through computational methods. One such method is agent-based simulations. Artificial agents can be placed in a simulated environment, equipped with information about similarities that hold between domains with stipulated properties, and instructed to draw analogical inferences about further similarities. Once again, varying the relevant conceptual uniformity levels of concepts across agents, and checking the success of the resulting inferences, would allow us to determine whether the proposed criterion is fertile, at least as a proof of concept. Another computational method that may be employed here is machine learning. We can train several competing neuro-symbolic models with analogical reasoning data. For example, we can train them with correct and incorrect analogies. The data should be such that the input features of different models correspond to concepts with different levels of relevant conceptual uniformity. Once the models are trained and tested on the existing data sets, we can unleash them on the world to see if any succeed, and indeed do better than the others, at making scientific discoveries.

It's important to note that, as far as we can see, no great obstacle stands in the way of computationally implementing the relevant conceptual uniformity admissibility criterion. For example, either via agent-based simulations or via a machine-learning neuro-symbolic model. If that is the case, and seeing as other aspects of analogical reasoning have already been successfully implemented *in silico*, it's safe to conclude that there are some grounds for hope in the assertion that analogical reasoning can be fully automated. Given the importance of analogical reasoning to heuristics, this, in turn, offers hope that scientific discovery, more generally, may one day be fully automated. Indeed, this is regardless of whether a universal approach to modelling analogical reasoning is feasible.

8. Conclusion

The foregoing discussion has, we hope, shed some light on analogical reasoning, its capabilities and limits. We began by exploring five major attempts (TSS, HCM,BAM, MPM, SME) at modelling analogical reasoning. Each of these made some headway towards that goal but also had some drawbacks. We then proceeded to outline an objection that affects all of them, namely Norton's argument that there is no universal model of analogical reasoning. We followed that up with an attempt to eliminate or at least reduce the objection's sting by positing an additional criterion for the goodness of analogical reasoning: the relevant conceptual uniformity admissibility criterion. We then questioned the satisfiability of this criterion with a challenge that has its roots in Wittgenstein's family resemblance metaphor. As a way of meeting the challenge, we argued that some concepts in natural science are indeed uniform, or, at least, more uniform than others, and that scientific inquiry strives towards such uniformity. We also tried to provide some initial motivation for the claim that analogical reasoning, in particular, and scientific discovery, more generally, can plausibly be automated.

One final thought is worth having. Bartha (2010) asks the very pertinent question "what reason do we have to expect analogical arguments to work?" and immediately responds that "[t]he best answer I can give is that our models of analogical reasoning provide a forum that lets us debate about, and ultimately identify, the 'right' critical factors, and hence the appropriate invariants for establishing symmetry between two domains" (303). This response leaves out the most important piece of the puzzle. The simple reason why analogical arguments work is because nature cooperates. Less metaphorically, nature contains considerable uniformity as well as repetition, and that is something that we can exploit. After all, bodies with mass attract each other, regardless of whether they are terrestrial or celestial. Masses and charges exhibit the same inverse square form of laws in the domains of gravitational and electrostatic phenomena. Selection and resource pressures determine survival and reproduction rates, irrespective of whether they were put in motion by the environment or by human hands. Newton's aphorism is highly instructive here: "Nature is after all simple, and is normally self-consistent throughout an immense variety of effects, by maintaining the same mode of operation" (Letter to Dr William Briggs, reproduced in Turnbull 1960: 418).

References:

Barrett, D., Hill, F., Santoro, A., Morcos, A., & Lillicrap, T. (2018, July). Measuring abstract reasoning in neural networks. In *International conference on machine learning* (pp. 511-520). PMLR.

Bartha, P. (2010). *By Analogical Reasoning*. Oxford: Oxford University Press.

Campbell, N.R. (1957). *Foundations of Science*. New York: Dover.

Currie, A. (2013). Convergence as evidence. *The British Journal for the Philosophy of Science*, 64, 763–86.

Darwin, C. ([1860] 1967). *Darwin and Henslow: the growth of an idea: letters 1831-1860*. University of California Press.

Drake, S. (1999). *Essays on Galileo and the History and Philosophy of Science* (Vol. 1). University of Toronto Press.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1986). The structure mapping engine. Proceedings of the American Association for Artificial Intelligence, AAAI, Philadelphia, PA: 272-277.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive science*, 7(2), 155-170.

Heering, P. (1992). On Coulomb's inverse square law. *American journal of physics*, 60(11), 988-994.

Hesse, M. B. (1966). *Models and Analogies in Science*. London: Sheed and Ward.

Himran, S., Suwono, A., & Mansoori, G. A. (1994). Characterization of alkanes and paraffin waxes for application as phase change energy storage medium. *Energy sources*, 16(1), 117-128.

Hofstadter, D. R. (1995). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. Basic books.

Hu, S., Ma, Y., Liu, X., Wei, Y., & Bai, S. (2021, May). Stratified rule-aware network for abstract visual reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 2, pp. 1567-1574).

Kuhn, T. S. (1961). The function of measurement in modern physical science. *Isis*, 52(2), 161-193.

Lange, M. (1993). Natural laws and the problem of provisos. *Erkenntnis*, 38(2), 233-248.

Lovett, A., & Forbus, K. (2017). Modeling visual problem solving as analogical reasoning. *Psychological review*, 124(1), 60.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, 85(3), 207.

Mill, J. S. ([1843]1973). *The Collected Works of John Stuart Mill (Books I-III)*. John M. Robson (Ed.), Toronto: University of Toronto Press.

Miller, D. (2017). *Out of error: Further essays on critical rationalism*. Routledge.

Mitchell, M. (2021). Abstraction and analogy-making in artificial intelligence. *Annals of the New York Academy of Sciences*, 1505(1), 79-101.

National Institute of Standards and Technology (NIST) 'Definitions of SI Base Units', https://www.nist.gov/si-redefinition/definitions-si-base-units

Newton, I. ([1687] 1999). *The Principia: mathematical principles of natural philosophy*. University of California Press.

Norton, J. D. (2003). A material theory of induction. *Philosophy of Science*, 70(4), 647-670.

Norton, J. D. (2011). History of science and the material theory of induction: Einstein's quanta, mercury's perihelion. *European Journal for Philosophy of Science*, 1, 3-27.

Norton, J. D. (2021). *The Material Theory of Induction*. University of Calgary Press.

Oppenheimer, R. (1956). Analogy in science. *American Psychologist*, 11(3), 127.

Reiss, J. (2015). A pragmatist theory of evidence. *Philosophy of Science*, 82(3), 341-362.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4), 573-605.

Rosenfeld, L. (1969). Newton's views on aether and gravitation. *Archive for history of exact sciences*, 6(1), 29-37.

Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Turnbull, H. W. (Ed.) (1960). *The Correspondence of Isaac Newton (vol. II)*. Cambridge: Cambridge University Press.

Webb, T. W., Sinha, I., & Cohen, J. D. (2020). Emergent symbols through binding in external memory. arXiv preprint arXiv:2012.14601.

Wittgenstein, L. (1953). *Philosophical Investigations*, translated by G. E. M. Anscombe, New York: Macmillan Publishing Co., Inc.

Zhang, C., Gao, F., Jia, B., Zhu, Y., & Zhu, S. C. (2019). Raven: A dataset for relational and analogical visual reasoning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5317-5327).