How Models of Creativity and Analogy Need to Answer the Tailorability Concern

John Licato, Selmer Bringsjord, Naveen Sundar Govindarajulu

1 Introduction

Creativity and the ability to reason analogically have a strong relationship [18, 19, 29, 23]. For example, many of the greatest lines in literature make use of familiar analogical processes: "A rose by any other name would smell as sweet" only makes sense if that rose and its name are understood to correspond to Romeo and his family name (analogical mapping). The subsequent analogical inference is that just as the nature of the rose is independent of its name, so is Romeo's nature independent of his.

Even if one believes that overuse of analogy can be harmful to creative thought [40], many researchers argue that the ability to determine analogical similarity is important at least for combinatorial creativity, which seems to be very easy for humans but very difficult for AI systems [5, 6]. Furthermore, the importance of analogy is of course not limited to creativity, as analogical ability has been identified as an indispensable component of artificial general intelligence as well [11, 27].

With the above in mind, it makes sense to develop models of analogy, both computational and theoretical. Much work has been done in this direction; a few implementations include SME [17], LISA [30], HDTP [26], and recently our own META-R [33]. On the surface, it seems that the current generation of analogical systems sufficiently capture and explain all of the phenomena commonly associated with analogical reasoning, and that they will eventually reach levels characteristic of human cognition. It may well be the case that the most important principles underlying the nature of analogy have been expressed. But a serious objection has been

John Licato
Rensselaer Polytechnic Institute, Troy, NY, e-mail: licatj@rpi.edu

Selmer Bringsjord
Rensselaer Polytechnic Institute, Troy, NY, e-mail: selmer@rpi.edu

Naveen Sundar Govindarajulu
Rensselaer Polytechnic Institute, Troy, NY, e-mail: govinn@rpi.edu

raised recently which, as will be argued, should be the primary focus of analogical researchers over the next few years — at least if any significant further progress is to be made in the direction of creativity and AGI.

The objection is raised by Gentner and Forbus (2011). They call it the 'tailorability concern' (TC), and the objection echoes a common criticism of cognitive systems in general: that they operate on toy examples manually constructed in such a way as to guarantee the desired solution. However, though this concern may have been stated in many forms throughout the years [37], it lacks, to our knowledge, a formulation clear enough to anchor productive scientific discussion. And this ambiguity in turn negatively impacts not only the relevant science, but AI engineering as well: absent a definition of TC, it is difficult to understand precisely what an analogical system must do in order to successfully answer TC. In the remainder, we take steps toward addressing this problem as it applies to analogical systems.

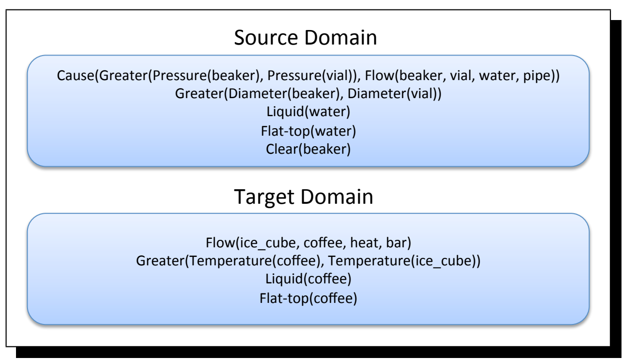

Fig. 1 One version of the popular "heat flow" example from Falkenhainer et al. (1989) used to test many analogical systems.

2 The Tailorability Concern

A frequently appearing criticism of cognitive systems in general is that they are only applied to manually constructed 'toy examples,' a problem many researchers in the field themselves acknowledge. Gentner and Forbus (2011) refer to the problem as the *tailorability concern* (TC): "that is, that (whether knowingly or not) the researchers have encoded the items in such a way as to give them the desired results" [24].

Of course, nothing is wrong with toy examples *per se*. They can be extremely useful in demonstrating key concepts, helping to illustrate particular qualitative strengths or weaknesses of computational models, or helping to get a new model off the ground. Indeed, the present authors plead guilty to using toy examples in these ways; but properly done, carefully chosen microcosmic cases can be very useful as demonstrations-of-concept. But we should be careful not to treat such examples as the final proof of a system's worth, since in most of these examples it is not clear that the principles used to solve them generalize to other problems, nor is it clear that such principles can be used to mechanically find useful solutions just as effectively in the absence of human assistance.

For example, the well-known water-flow/heat-flow analogy (Figure 1) has been used as a demonstration of many models of analogy [17, 28, 42, 32]. But little to nothing is written about how the structural representations used in examples such as these are acquired in the first place. One approach is to model the acquisition of structured representations through sensory data (e.g., see [16]), and another is to presume the existence of a large database of already-structured data (such as that to which a neurobiologically normal adult might be expected to have access), and some sort of filtering process that ensures that from this database, proper representations are selected and any unnecessary data that would produce incorrect matching results are excluded.[1] Yet even when such filtering processes are proposed, they are not put to the test and proven to perform well with a large database containing enough knowledge to match that of a child's, much less an adult's. The TC rightfully attempts to refocus efforts on these filtering processes, by requiring that they demonstrate the ability to produce clean source and target analogs as required by the analogical mappers.

The field of case-based reasoning (CBR), which overlaps quite heavily with that of analogical reasoning (AR), also deals with some of the issues raised by the TC. There are differing opinions on what features distinguish the CBR and AR approaches (see [1, 14, 35]), but two common themes are that CBR tends to deal with source and target cases that come from the same domain, and cases are selected and adapted with some clear pragmatic goal in mind. AR approaches, on the other hand, try to be more generally applicable across different domains, and tend to focus more on the mapping process that actually determines analogical similarity. CBR approaches, then, deal with the TC by trading generality for effectiveness, so that a program designed to work well in one domain (medical diagnosis, for example, is a popular field) may not work so well in another without a significant amount of human assistance.

Unfortunately, the CBR approach of restricting generality does not sufficiently answer the TC. Analogy research can be seen as centering around fundamental questions, one of them being: How can we find good analogies? The TC is especially problematic because it forces analogy researchers to prove that their theoretical process is the answer to this question, and although it can be promising to see that some particular approach produces good analogies in some limited domain, no approach can constitute a completely satisfying answer to this question unless it is versatile enough to perform well in *many* domains.

---

[1] It would be less appropriate to urge a careful treatment of the TC and tie it so closely to large semantic databases if they weren't available. But over the past few years, natural-language processing and semantic-web technologies have been progressing to the point where we now have access to large collections of semantic databases containing wide-ranging general knowledge. These include Cyc [36], Freebase [7], and DBPedia [3]. Many of these have easy-to-use interfaces.

Any system that can answer the challenge of the TC will instantly distinguish itself from every other extant analogical system, since (at least, to our knowledge) the only one that has been able to do this with some degree of success is the SME-based family of systems [24, 34, 20]. For this reason, we should clarify what it means to answer this challenge and discuss why it is such a non-trivial feat.

2.1 Answering the TC

Gentner and Forbus (2011) suggest that there are two possible ways to answer the TC. One applies to visual domains, and involves using automatic encodings of visual representations. The other more generally applicable direction involves two key features: first, the use of pre-existing databases; second, an automated or semi-automated parsing process that goes from input text to a sufficiently rich semantic representation. A first attempt at a precise statement of what it means to answer the TC is as follows:

> **TCA$_1$** A computational system of analogy answers the TC if, given no more than a pre-existing database and an unparsed input text, it is able to consistently produce *good* analogies across many domains.

At least one general intent behind **TCA$_1$** is clear: it attempts to place emphasis on the filtering process (whose job is, as we said, to either select some subset of available source analogs from a large database and recommend only some of them for the more computationally expensive step of analogical matching, or to automatically construct structured representations from sensory data). By removing the reliance on human intervention, TC ensures that the filtering is not done manually in such a way that guarantees desired results. However, in order to truly answer the TC in a satisfactory way, we must be precise about its purpose and motivations: What concerns are behind the TC in the first place? Furthermore, **TCA$_1$** is hopelessly vague: the words ‘unparsed’ and ‘good,’ if left open to interpretation, make it too easy for anyone to prematurely claim victory over TC. Also: Why is it important that the database be pre-existing? What degree of separation must there be between the creators of the database and the designers of the analogical system? For example, does the database’s underlying knowledge-representation philosophy need to overlap with that of the analogical system?

2.1.1 What is a ‘Good’ Analogy?

Though the widely influential Structure-Mapping Theory [22] offers the systematicity principle and the one-to-one constraint as indicative features of a good analogy, it does not provide a clear quantitative measure of analogy quality. SME evaluates match scores by combining the scores of the evidence provided by its match rules; this allows for a comparison between different matches of the same problem. But the resulting match score is not normalized, and as a result match quality between different problems cannot be compared [21]. Other analogical models do not help much in this regard. Holyoak and Thagard's (1989) Multiconstraint Theory, for example, introduces additional criteria to evaluate what makes an analogical match a good one, making cross-domain analogical match quality more difficult to assess.

This is especially problematic when considering TC. Given a set of predefined problems and desired answers for all of them, the good analogy is simply the one that performs as intended. But when applied to problem sets where the best analogy, or even the existence of a good analogy, is not clear even to the persons running the experiment, do the guidelines currently available still apply?

Paul Bartha (2010) offers a comprehensive theory of analogy that can be useful here. His goal is to produce a *normative* account of analogy, and in the process he sheds light on the proper role of analogy within a larger context of reasoning. His Articulation Model is based on the idea that there are two features common to all good analogical arguments: a prior association (a relationship in the source analog that is transferred to the target), and a potential for generalization. But perhaps most relevant to our present discussion is his claim that by virtue of the analogical argument, the resulting hypothesis inferred through an analogical argument contains no more than *prima facie* plausibility, which can be understood as something like a suggestion that the inferred hypothesis is worth exploring further. If there is an independent reason to reject the hypothesis, such as a deductive argument showing it leads to a contradiction, or contrary empirical evidence, then it can be abandoned.

The idea that the proper role of an analogical argument is to do no more than provide *prima facie* plausible hypotheses [39, 38] suggests that the relationship between an analogy's match quality and its tendency to produce hypotheses which can be independently verified may not be as simple as it might seem. In the end, a *model* of analogy is a good one only if it produces good analogies, and an analogy is a good one only if it produces plausible hypotheses.[2]

This complicates things further. In toy examples, the information available to the analogical matcher is very limited; a hypothesis in these cases is plausible if the information available to the analogical matcher does not provide a reason to reject that hypothesis. But if we are suggesting that the information available to the analogical system (or more specifically, the filtering process) is actually a large database conceivably encompassing, say, all of the knowledge on the World Wide Web, then how are we to realistically evaluate the system's performance?

In light of these complications, we propose the following alterations of $\mathbf{TCA_1}$. If the word 'good' is replaced with 'useful,' then we connote an evaluation method that is not based in the particular metrics as defined by the analogical theory itself (which can be controversial), but rather based in an intuitive notion that can and should be evaluated independently of the model. In other words, researchers might disagree on how to measure an analogical match, but whether the resulting analogically in-

[2] We might leave room here to exclude models of analogy that have psychological or neurological plausibility as their primary end goals. In these cases, it might be the goal of the model to replicate poor analogical reasoning as well, if it matches human performance. But it is our assumption (at least in the present inquiry) that the ultimate goal of AGI research is not to model poor human reasoning.

ferred hypothesis is useful can be evaluated without any knowledge of the analogical matcher used. Of course, since what is 'useful' can be very domain-dependent, we do not claim that this word is completely unambiguous. Later in this paper, a better replacement for the word 'useful' will be suggested. For now, the aim behind this move is to divorce the metric which determines the quality of an analogical match's results (which may be very domain-dependent) from the theory-specific metric that the matcher is specifically designed to optimize. That (deceptively) small change gives us $\mathbf{TCA_2}$:

> $\mathbf{TCA_2}$ A computational system of analogy answers the TC if, given no more than a pre-existing database and an unparsed input text, it is able to consistently produce *useful* analogies across many domains.

2.1.2 What are Acceptable Databases?

In $\mathbf{TCA_2}$, it is clear that the knowledge available in the database used is a limiting factor in how informative the inferences produced by the analogical system can be. The suggestion phrased by Gentner and Forbus (2011) as "pre-existing databases" requires more clarification. The implication (at least as we interpret it) is that the dataset and the structures within cannot have been constructed for the purpose of solving the particular toy examples that are of interest. Otherwise this introduces bias and tailorability concerns, in spite of the best intentions of the designers. Two issues immediately come to mind. First, what is the proper level of separation between the database and the analogical system? Secondly, how large does the database have to be?

The first question is at least partially answered by considering the level of *representational agreement* between the database and the analogical system. For example, if the database is a purely distributed one with no localist concepts whatsoever (which is, we acknowledge, an unlikely possibility), and the analogical system is one that uses only localist, explicitly structured data, then a significant level of work will be needed to first extract the information from the database and put it into the form that the analogical reasoner requires (this can be considered a *re-representation* step [15, 43, 31]). The choice of database becomes important for this reason, and if no database exists that does not require a high level of re-representation, then it suggests a problem: Consider that although proponents of localist, distributed, and hybrid representation styles make claims all the time about the scalability of their assumptions of knowledge representation, the researchers who have to design and work with large semantic databases actually have to put these ideas to the test. If the state-of-the-art research in the database field has significant difficulty with some style of representation as required by an analogical matcher, then perhaps that matcher needs to carefully consider its assumptions about the nature of knowledge representation, or else be able to easily extract the necessary data.[3] The choice of database is therefore a sensitive one.

[3] We do not mean here to say that what works best for large artificial databases is the same as what is employed by the human brain. But if a researcher declares that the structure of human knowledge

It may help to go back to the criticisms which may have motivated this requirement in the first place. Hofstadter's group criticizes SME by saying that "the analogy is already effectively given in the representations" [37]. The information provided to the matcher is selected in such a way that it does not include a significant amount of extra data that would lead to false analogies, and the choice of predicates and objects is done in such a way that presupposes the desired matching. This is a criticism composed of two points: the amount of information provided as input (which is not too little, or too much, but *just right*), and the nature of the information (the corresponding predicates on both sides just happen to have the right amount of arguments in the necessary order). The system is not robust enough to produce good analogies when given the same input data in a variety of formats.

Two possible changes can begin to answer these critiques. One is to require that the databases are large enough, and the input minimal enough, to introduce a significant amount of false matches that would confuse less robust analogical matchers. Critics claiming that the input is too carefully chosen could be pointed to the fact that the analogical matcher must plumb through an answer space that is large (at least relative to the input problem). The larger the search space, the more impressive the ability of the analogical matcher to select only a few potentially relevant source analogs. Furthermore, an inability to demonstrate scalability to large datasets weakens any architecture's claim to psychological plausibility: if the architecture can't handle a dataset large enough to produce non-trivial answers, how can it be an accurate model of a process used by human-level reasoners?[4]

Secondly, we could require robust and consistent performance on a variety of input forms. For example, in the heat-flow problem (Figure 1) Mitchell and Hofstadter (1995) note that there are many possible ways to structure the input: heat could be described as an object, or as an attribute of coffee, or *heat flow* could be a relation with three rather than four arguments [37]. Consistent performance across various input forms puts more pressure on the analogical matcher's re-representation algorithm(s), rather than relying on a separate NLP module. This also allows for a leveling of the playing field across different systems: In order to show that a given example adheres to this requirement, a localist, structured analogical system would have to demonstrate two things with regard to that particular example:

- Across multiple, more-or-less equivalent structural representations of the same input data, and a wide variety of domains, the matcher still produces the desired results.
- The desired results are still produced when the input is minimal; meaning any redundant information or structural constructs which might be identified by critics as being used only to aid the matcher can be removed.

Of course, if a system did happen to require no more than unstructured, natural-language descriptions as input, or direct sensory data from visual representations, it would satisfy both of these conditions. This allows our criteria to encompass the alternate route to answering the TC mentioned by Gentner and Forbus (2011)—a

has certain properties, and large datasets cannot be created that do not have those properties for reasons of scalability, then it should be at least a weak hint that perhaps the assumption of those properties is not practicable.

[4] This is a common criticism of Hummel and Holyoak's LISA system; see [24].

route which seeks to answer TC by not necessarily having a large database, but by having one that at least attempts to directly construct structured representations from low-level sensory or natural-language data. The generality of these conditions allows us to claim the converse of $\mathbf{TCA_2}$, leading us to our next iteration:

$\mathbf{TCA_3}$ A computational system of analogy answers the TC if and only if given no more than either

- unstructured textural and/or visual data, or
- a large, pre-existing database,

and minimal input, it is able to consistently produce *useful* analogies and demonstrate stability through a variety of input forms and domains.

One might be satisfied with this set of criteria, which draws its strength from its lack of commitment to any particular theory of analogy, and its emphasis on large sets of non-tailored input. But $\mathbf{TCA_3}$ is undeniably ambiguous, and may not be focused enough to guide any research program. We encourage the reader to take $\mathbf{TCA_3}$ and develop it further, but first, to close out this paper we will take some steps of our own to reduce some of its weaknesses.

2.1.3 Strengthening the TC with Psychometric AI

We will make two important moves to sharpen $\mathbf{TCA_3}$. One, we turn to *Psychometric AI* (PAI) [9, 13], according to which, in a nutshell, commendable AI systems are those that demonstrate prowess on tests of various mental abilities from psychometrics. Our second move is to imbed analogy-generation systems within broader AI problem-solving systems that make use of additional forms of fundamental reasoning in human-level intelligence; e.g., deduction. In particular, we place TC within the context of the integration of analogical reasoning with deduction, which we dub *analogico-deductive reasoning* (ADR), and which we have explained and demonstrated elsewhere [11]. An ADR system does use analogy generation, but analogies are used to guide solutions that can be rigorously verified by proof. The architecture-sketch of an ADR system that accords with our pair of moves is shown in Figure 2. While we don't have the space to provide details here, this system receives puzzles that are part linguistic and part visual in nature (e.g., so-called *seating puzzles*), and harnesses not only our own analogical system (META-R), and not only ATP technology, but *visual* theorem-proving; for formal details see [2]. This architecture-sketch is inspired by, but abstracts and extends beyond, AI systems able to solve analogy-relevant problems. One such inspiring system is by Lovett, Forbus, and Usher [34]; it can solve items on the Raven's Progressive Matrices. But it cannot for example prove that its answers are correct, which is part of what the architecture-sketch in Figure 2 demands.

Given the context we have now set, we can articulate a new biconditional:

$\mathbf{TCA_4}$ A computational system $\mathscr{A}$ for analogy generation answers the TC if and only if, given as input no more than either

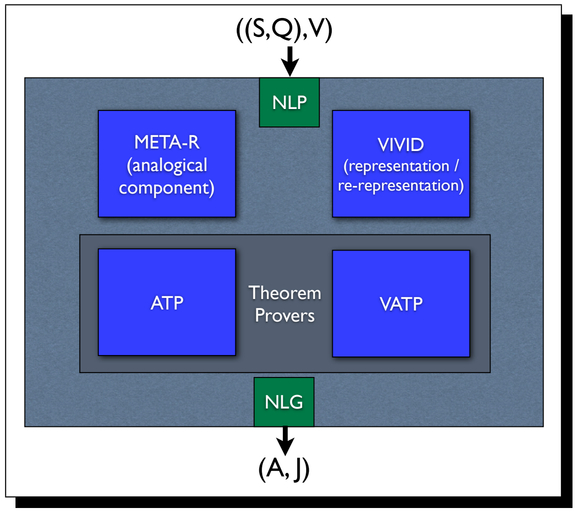

Fig. 2 Architecture-Sketch of an ADR System That Answers TC. Input includes a **S**tory and a **Q**uestion, along with **V**isual content. Output includes an **A**nswer and—in the form of a proof or at least a rigorous argument—a **J**ustification.

- unstructured textual and/or visual data, or
- a vast, pre-existing database not significantly pre-engineered ahead of time by humans for any particular tests of $\mathscr{A}$,

is — in keeping with aforementioned *Psychometric AI* — able to consistently generate analogies that enable $\mathscr{A}$ to perform *provably well* on precisely defined tests of cognitive ability and skill.

To comment briefly on $\mathbf{TCA_4}$, first note that we remove the explicit requirement that the ability to find useful analogies be stable across a variety of input forms and domains. This is subsumed by the requirement of good performance on precisely defined tests; it is assumed that a sufficiently difficult psychometric test would provide questions that are both varied in their form (e.g., word problems, puzzle solving, story comprehension) and in their domains. The implicit requirement of domain variety rules out the possibility of an artificial reasoning agent that can only process, for example, certain types of math problems, as an acceptable answer to the TC.

Some might also see $\mathbf{TCA_4}$ and the use of PAI as too restrictive in that it relies too heavily on problem-solving and not enough on either creative thinking or the kind of everyday analogical thinking that may not be goal-oriented in nature.[5] PAI, however, provides a tool for measuring those abilities that, at least at the surface, don't rely on directed problem-solving, such as reading comprehension. Additionally, it is difficult to imagine that any research program in AGI would be able to demonstrate clear progress without showing increased performance in an ability that can be measured according to some psychometric test. Non-goal-oriented analogical reasoning

[5] We thank our reviewers for their insightful comments on this topic and others.

is a good example of this principle: If the cognitive processes underlying normal analogical reasoning when it is non-goal-oriented (as in everyday reasoning) and when it is goal-oriented (as during psychometric testing) are largely the same, then an artificial system capable of performing the latter has a strong claim to performing the former. A system that only has sporadic performance on psychometrically-measurable tasks is difficult to defend as generally intelligent.

One might ask: Can a system do well on psychometric tests and still be subject to claims of tailorability? The answer, if the requirements in $\mathbf{TCA_4}$ are not met, is certainly *Yes*. PAI is not meant to be a replacement for the input format and large database requirements we have been developing in this paper; rather, it is only one possible sharpening of the ambiguous concepts in $\mathbf{TCA_3}$. Other possibilities may exist, but we do not at present know of any satisfying alternatives.

3 Case Mapper: A Step in the Right Direction

$\mathbf{TCA_3}$ and $\mathbf{TCA_4}$ reflect a growing realization among the AI community that if some cognitive system is to ever move closer to AGI, it needs to be able to demonstrate its performance with a large knowledge base. *Case Mapper*, from Northwestern University's Qualitative Reasoning Group, is a system currently in development that allows users to test SME [17], MAC/FAC, and other analogical reasoning tools, along with large databases that are included. Among these databases is a version of OpenCyc [36], which comes with millions of facts connecting tens of thousands of concepts and relations. Although in our opinion the database is not yet large enough to match the performance of an 8-year-old child on analogy problems that require recollection of arbitrary source cases, the way it gives the user access to such a collection of powerful tools suggests it can become very useful to anyone interested in the field.

Case Mapper works with *cases*, which are collections of facts that can be constructed automatically by the program given some concept in the database. After mapping two cases, a set of candidate inferences can be drawn from the source to the target case. We can explore these candidate inferences manually in order to have some idea of the system's current ability. For example, having Case Mapper construct cases for *Dog* and *Cat* and setting them as source and target cases respectively, among the candidate inferences produced are:

```
(Collection
  (SubcollectionOfWithRelationToTypeFn
    (:skolem Shedding)
    (:skolem bodilyDoer)
    Cat
  )
)
```

This essentially means, "Cats shed," which is an hypothesis that could easily be confirmed by observation, allowing an ADR-based system to add to its own knowledge base autonomously. Turning now back to $\mathbf{TCA_4}$, we note that a similar approach can be used to solve Psychometric-styled analogy questions,[6] such as the following:

Baseball-Ball : BaseballGame :: TennisBall : ?

In order to solve a problem like this, a system might find a mapping between a source case constructed around the concepts *Baseball-Ball* and *BaseballGame*, and a target case from *TennisBall*, where the mapping is instructed beforehand to map the two ball concepts together. In the resulting mapping, whichever element in the target case corresponds to *BaseballGame* should be the correct answer. All of these settings are available in the current version of Case Mapper, but it does not output the correct answer because there is not enough information in the cases to draw a mapping with confidence. The relevant facts in the provided OpenCyc database are as follows:

```
(isa
  Baseball-Ball
  (EquipmentTypeForEventTypeFn
    BaseballGame
  )
)

(isa
  TennisBall
  (EquipmentTypeForStructuredActivityFn
    Tennis-TheGame
  )
)
```

Note that although a human would be expected to realize that in these contexts the symbols *EquipmentTypeForEventTypeFn* and *EquipmentTypeForStructuredActivityFn* are close together enough semantically to map for the purposes of this problem, they do not register as a match here, and as a result the concepts *BaseballGame* and *Tennis-TheGame* are not linked. Nevertheless, were the database more complete (as it will likely be in the future), such a psychometric problem would have been trivial.

[6] We must mention here that standard proportion-type analogy questions seem to be falling out of favor recently; they have for example been removed from the SAT, and it is unclear at this time whether they will be brought back anytime soon.

4 Conclusion and Future Work

The concerns reflected by TC have been more or less echoed in many criticisms of analogical systems over the past 30 years or so, but little to no discussion has attempted to precisely state the conditions under which an analogical system would no longer be subject to this common complaint. We humbly acknowledge the limitations of our formulation here, and remind the reader that even a system with a strong claim to meeting the conditions of $\mathbf{TCA_4}$ (e.g., an implementation of Figure 2) may not be completely immune to TC-based criticisms. But it seems that such criticisms would be somewhat difficult to justify. We hope that this paper inspires further discussion, and invite researchers to pick up from and sharpen $\mathbf{TCA_4}$, or to steer away from Psychometric AI and therefore develop $\mathbf{TCA_3}$ in a different direction.

Those who strive for something deeper than a test-based anchor for a response to TC will presumably be more sanguine than Bringsjord about the ultimate reach of computing machines. Bringsjord holds that it is impossible for any mere computing machine to *originate* anything [10], or even to genuinely understand anything [12]. But he holds on the other hand that for any determinate behavioral test, a computing machine can be programmed to pass it—even if the test is one which, if passed, suggests to most human observers that the passing machine is not simply the product of human intelligence, harnessed [8]. In short, the turn to $\mathbf{TCA_4}$ as a way of fleshing out $\mathbf{TCA_3}$ reflects deep pessimism about computing machines, and specifically a deep affirmation of the penetrating and profound nature of TC. Licato, in contrast, is more optimistic about the potential of analogical-hybrid systems, but is not prepared to take any philosophical positions on whether a system that passes TC demonstrates true understanding or creativity.

What about AI work devoted to building creative systems on the strength of analogical reasoning? Our work continues along this path. We are currently investigating the potential of using our analogical matching system META-R, which is designed to allow for a flexible application of heuristics across a wide variety of domains, with the community-built dataset Freebase [7]. We are also investigating the use of META-R in modeling the automatic generation of logical theorems and high-level proofs from semi-formal and formal domains [33, 25]. Future work in this direction includes the investigation of Analogico-Deductive Moral Reasoning (ADMR) using a hybrid of CLARION [41] and LISA [30]. And finally, we are seeking to leverage analogical reasoning in order to engineer systems capable of automatic programming.[7]

[7] In the case of automatic programming, the input shown in Figure 2 would be instantiated as the informal definition of a number-theoretic function (where that definition can be partly linguistic and partly visual), and the answer is code in some conventional programming language, accompanied by a proof of the correctness of this code relative to the input. Automatic-programming systems seemingly require the ability to judge two programs analogous. More precisely, such systems seemingly would need to be able to answer this question: Given a set of programs $\mathscr{P}$ in some programming language, can the system produce a similarity metric $\rho: \mathscr{P} \times \mathscr{P} \rightarrow \mathbb{R}$ capturing which pairs of programs are semantically analogous?

With profound gratitude, the support of both the John Templeton Foundation and AFOSR is acknowledged.

References

1. Aamodt, A., Plaza, E.: Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications **7**(1), 39–59 (1994)
2. Arkoudas, K., Bringsjord, S.: Vivid: An AI Framework for Heterogeneous Problem Solving. Artificial Intelligence **173**(15), 1367–1405 (2009). URL `http://kryten.mm.rpi.edu/vivid_030205.pdf`. The url http://kryten.mm.rpi.edu/vivid/vivid.pdf provides a preprint of the penultimate draft only. If for some reason it is not working, please contact either author directly by email.
3. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: Proceedings of the 6th International Semantic Web Conference (ISWC2007) (2007)
4. Bartha, P.F.: By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments. Oxford University Press (2010)
5. Boden, M.A.: Creativity and Unpredictability. Stanford Humanities Review **4**(2), 123–139 (1995)
6. Boden, M.A.: Computer Models of Creativity. AI Magazine **30**(3) (2009)
7. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. In: Proceedings of the 2008 ACM SIGMOD International conference on Management of data (SIGMOD '08). ACM (2008)
8. Bringsjord, S.: Could, How Could We Tell If, and Why Should–Androids Have Inner Lives? In: K. Ford, C. Glymour, P. Hayes (eds.) Android Epistemology, pp. 93–122. MIT Press, Cambridge, MA (1995)
9. Bringsjord, S.: Psychometric Artificial Intelligence. Journal of Experimental and Theoretical Artificial Intelligence **23**(3), 271–277 (2011)
10. Bringsjord, S., Ferrucci, D., Bello, P.: Creativity, the Turing Test, and the (Better) Lovelace Test. Minds and Machines **11**, 3–27 (2001)
11. Bringsjord, S., Licato, J.: Psychometric Artificial General Intelligence: The Piaget-MacGyver Room. In: P. Wang, B. Goertzel (eds.) Theoretical Foundations of Artificial General Intelligence. Atlantis Press (2012). URL `http://kryten.mm.rpi.edu/Bringsjord_Licato_PAGI_071512.pdf`
12. Bringsjord, S., Noel, R.: Real Robots and the Missing Thought Experiment in the Chinese Room Dialectic. In: J. Preston, M. Bishop (eds.) Views into the Chinese Room: New Essays on Searle and Artificial Intelligence, pp. 144–166. Oxford University Press, Oxford, UK (2002)
13. Bringsjord, S., Schimanski, B.: What is Artificial Intelligence? Psychometric AI as an Answer. In: Proceedings of the 18^{th} International Joint Conference on Artificial Intelligence (IJCAI–03), pp. 887–893. Morgan Kaufmann, San Francisco, CA (2003). URL `http://kryten.mm.rpi.edu/scb.bs.pai.ijcai03.pdf`
14. Burstein, M.H.: Analogy vs CBR: The Purpose of Mapping. In: Proceedings from the Case-based Reasoning Workshop, pp. 133–136. Pensacola Beach, Florida (1989)
15. Chi, M.T., Feltovich, P.J., Glaser, R.: Categorization and Representation of Physics Problems by Experts and Novices. Cognitive Science **5** (1981)
16. Doumas, L.A., Hummel, J.E., Sandhofer, C.: A Theory of the Discovery and Predication of Relational Concepts. Psychological Review **115**, 1–43 (2008)
17. Falkenhainer, B., Forbus, K.D., Gentner, D.: The Structure-Mapping Engine: Algorithm and Examples. Artificial Intelligence **41**, 1–63 (1989)
18. Finke, R.: Creative Imagery: Discoveries and Inventions in Visualization. Lawrence Erlbaum Associates (1990)

19. Finke, R., Ward, T., Smith, S.: Creative Cognition: Theory, Research, and Applications. MIT Press, Cambridge, MA (1992)
20. Forbus, K., Klenk, M., Hinrichs, T.: Compaion Cognitive Systems: Design Goals and Lessons Learned So Far. IEEE Intelligent Systems **24**(4), 36–46 (2009)
21. Forbus, K.D., Gentner, D.: Structural Evaluation of Analogies: What Counts. In: Proceedings of the Eleventh Annual Conference of the Cognitive Science Society, vol. 34 (1989)
22. Gentner, D.: Structure-Mapping: A Theoretical Framework for Analogy. Cognitive Science **7**, 155–170 (1983)
23. Gentner, D., Brem, S., Ferguson, R., Wolff, P., Markman, A., Forbus, K.: Analogy and Creativity in the Works of Johannes Kepler. In: T. Ward, S. Smith, J. Vaid (eds.) Creative Thought: An Investigation of Conceptual Structures and Processes, pp. 403–459. American Psychological Association (1997)
24. Gentner, D., Forbus, K.: Computational Models of Analogy. Wiley Interdisciplinary Reviews: Cognitive Science **2**(3), 266–276 (2011)
25. Govindarajulu, N.S., Licato, J., Bringsjord, S.: Small Steps Toward Hypercomputation via Infinitary Machine Proof Verification and Proof Generation. In: Unconventional Computation and Natural Computation - 12th International Conference (UCNC 2013) (2013)
26. Gust, H., Kühnberger, K.u., Schmid, U.: Metaphors and Heuristic-Driven Theory Projection (HDTP). Theoretical Computer Science **354**(1), 98–117 (2006)
27. Hofstadter, D.R.: Epilogue: Analogy as the Core of Cognition. In: D. Gentner, K.J. Holyoak, B.N. Kokinov (eds.) The Analogical Mind: Perspectives from Cognitive Science, chap. 15. The MIT Press (2001)
28. Holyoak, K.J., Thagard, P.: Analogical Mapping by Constraint Satisfaction. Cognitive Science **13**(3) (1989)
29. Holyoak, K.J., Thagard, P.: Mental Leaps: Analogy in Creative Thought. MIT Press, Cambridge, MA (1995)
30. Hummel, J.E., Holyoak, K.J.: A Symbolic-Connectionist Theory of Relational Inference and Generalization. Psychological Review **110**, 220–264 (2003)
31. Krumnack, U., Gust, H., Kühnberger, K.u., Schwering, A.: Re-Representation in a Logic-Based Model for Analogy-Making. In: Proceedings of the 21st Australasian Joint Conference on Artificial Intelligence. Aukland, New Zealand (2008)
32. Krumnack, U., Schwering, A., Gust, H., Kühnberger, K.u.: Restricted Higher-Order Anti-Unification for Analogy Making. In: Proceedings of Twenties Australian Joint Conference on Artificial Intelligence, LNAI 4830, pp. 273–282. Springer (2007)
33. Licato, J., Govindarajulu, N.S., Bringsjord, S., Pomeranz, M., Gittelson, L.: Analogico-Deductive Generation of Gödel's First Incompleteness Theorem from the Liar Paradox. Proceedings of the 23rd Annual International Joint Conference on Artificial Intelligence (IJCAI-13) (2013)
34. Lovett, A., Forbus, K., Usher, J.: A Structure Mapping Model of Raven's Progressive Matrices. In: Proceedings of CogSci, vol. 10 (2010)
35. Mantaras, R.L.d., Mcsherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M.L., Cox, M.T., Forbus, K., Keane, M., Aamodt, A., Watson, I.: Retrieval, Reuse, Revision and Retention in Case-based Reasoning. The Knowledge Engineering Review **20**(3), 215–240 (2005)
36. Matuszek, C., Cabral, J., Witbrock, M., DeOliveira, J.: An Introduction to the Syntax and Content of Cyc. In: Proceedings of the 2006 AAAI Spring Symposium on Formalizing and Compiling Background Knowledge and its Applications to Knowledge Representation and Question Answering, vol. 3864 (2006)
37. Mitchell, M., Hofstadter, D.R.: Perspectives on Copycat: Comparisons with Recent Work. In: D.R. Hofstadter, the Fluid Analogies Research Group (eds.) Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought, chap. 6. Basic Books, Inc (1995)
38. Norton, J.D.: Formal and Material Approaches to Analogical Inference. Forthcoming (Forthcoming)

39. Owen, S.: Analogy for Automated Reasoning. Academic Press (1990)
40. Riesbeck, C.K., Schank, R.C.: Inside Case-based Reasoning. Lawrence Erlbaum Associates (1989)
41. Sun, R.: Duality of the Mind: A Bottom Up Approach Toward Cognition. Lawrence Erlbaum Associates, Mah- wah, NJ (2002)
42. Wilson, W.H., Halford, G.S., Gray, B., Phillips, S.: The STAR-2 Model for Mapping Hierarchically Structured Analogs. In: D. Gentner, K.J. Holyoak, K.K. Boicho (eds.) The Analogical Mind: Perspectives from Cognitive Science, chap. 4. The MIT Press, Cambridge, MA (2001)
43. Yan, J., Forbus, K., Gentner, D.: A Theory of Rerepresentation in Analogical Matching. In: Proceedings of the 25th Annual Conference of the Cognitive Science Society (2003)