Qualitative Mechanical Problem-Solving by Artificial Agents: 
Further Progress, Under Psychometric AI

Shreya Banerjee (shreyaabannerjee@gmail.com) 
Selmer Bringsjord (Selmer.Bringsjord@gmail.com) 
Michael Giancola (mike.j.giancola@gmail.com) 
Naveen Sundar Govindarajulu (Naveen.Sundar.G@gmail.com) 

Rensselaer AI & Reasoning (RAIR) Lab 
Department of Computer Science, Department of Cognitive Science 
Rensselaer Polytechnic Institute (RPI), 110 8th Street 
Troy NY 12180 USA

Abstract

Qualitative mechanical problem-solving (QMPS) is central to human-level intelligence. Human agents use their capacity for such problem-solving to succeed in tasks as routine as opening the tap to drink or hanging a picture on the wall, as well as for more sophisticated tasks in demanding jobs in today’s economy (e.g., emergency medicine, plumbing, hydraulic machinery, & driving). Unfortunately, artificial agents (including specifically robots) of today lack the capacity in question. Our work takes QMPS to fall under the general, longstanding AI area of qualitative reasoning (QR), historically an intensely logic-based affair. We embrace this history, and take new, further steps to advance QMPS. The Bennett Mechanical Comprehension Tests (BMCT-I and BMCT-II) assess a human’s ability to solve QMPS problems, and are used in the real world by many employers to evaluate job candidates. Building on the work of others who have attacked BMCT under the rubric of Psychometric AI (PAI), we introduce one of our novel algorithms ($\mathcal{A}_{1B}$) in a family ($\mathcal{A}_B$) of such for QMPS as required by BMCT, illustrate via case studies, report time-based performance of $\mathcal{A}_{1B}$, and assess our progress with an eye to future work in which our approach is extended to a subclass of algorithms in $\mathcal{A}_B$ that exploit the power of argument-based nonmonotonic logic, and leverage the success of transformer models to enhance their efficiency.

Introduction

Qualitative problems differ from quantitative problems in that the former can be solved with scant discipline-specific information and without any robust quantitative calculation. For example, in the seminal work on qualitative physics by Hayes (Hayes 1978; 1985), qualitative reasoning (QR) about the behavior of fluids exploits not physics proper, but principles present in general human knowledge. Qualitative mechanical problem-solving (QMPS), in which problem-solving is based upon QR, is central to human-level intelligence. Humans use their capacity for such problem-solving to succeed on tasks as routine as opening the tap to drink or hanging a picture on a wall, as well as for more sophisticated tasks in demanding jobs in today’s economy (e.g., in emergency medicine, plumbing, & hydraulic machinery). Unfortunately, AIs and robots of today lack the capacity in question. We seek to contribute toward a time when artificial agents too can solve important classes of qualitative mechanical problems using advanced reasoning techniques. But what should guide research intended to make such a contribution? Psychometric AI (PAI) (Bringsjord and Schimanski 2003) provides an answer; it maintains that AI work should be carried out in pursuit of success on well-defined, longstanding, and highly regarded tests of the capability in artificial agents that is sought. For capability specifically in QMPS, (Klenk et al. 2011) avowedly adopted a PAI orientation, by working toward an AI capable of succeeding on the Bennett Mechanical Comprehension Tests (BMCT-I and BMCT-II) (Bennett 2008). The Bennett, or simply BMCT (as we shall hereafter call the pair -I and -II tests in tandem), has long been used to assess human ability to solve mechanical problems, and is used by many employers to evaluate job candidates. (Klenk et al. 2011) is based upon the power of analogy-making to tackle Bennett problems, and explicitly leaves aside formal, logicist approaches (a point expanded upon in the next section). We herein report on a formal approach to the Bennett.

The following section further describes QMPS, BMCT, and some related prior work. Following that, we illustrate our approach and subsequently analyze the implementation of our first novel algorithm for QMPS in BMCT.

Background and Related Work

Quantitative versus qualitative problem solving can be described as below:

- **Quantitative Problem Solving**: Solving a problem based on the actual quantity of attributes (value of length, volume, distance, etc) as employed in the particular sciences (e.g. physics) rather than using its quality for comparative measure.

- **Qualitative Problem Solving**: Solving a problem based on the quality of attributes (comparison of size, appearance, symmetry,
volume, length, etc.) as employed in everyday cognition rather than its actual quantitative measure.

Qualitative mechanical problem solving (QMPS) refers to solving qualitative problems by QR in the mechanical domain. This is of course an enormous domain. Hence, we restrict focus in it in accordance with Psychometric AI (PAI) (Bringsjord and Schimanski 2003), according to which daunting goals for AI are rigorized and restricted to corresponding tests of cognitive power at the human-level. Under this rubric in the present case (QMPS) we focus on BMCT. This specific move is taken and followed by (Klenk et al. 2011), who attack BMCT as a tractable, focused way to attack QMPS.

The Bennett measures a human-level agent’s ability to solve mechanical problems in practical situations. Many occupations require viable job candidates to have a certain level of understanding about application of mechanical principles, and BMCT is often used to evaluate candidates in this regard. Problems consist of diagram/s with labels, markings and textual information, a separate or combined — to now use psychometric terminology — story and question (= stem), multiple choice question (MCQ) with number of options ranging from 3–5, where each stem has only one correct option (i.e. key). See Figure 1. (Klenk et al. 2011) invent and demonstrate an excellent approach to a proper subset of BMCT problems, using analogical model formulation with sketches using their sketch system (sKEA). However, as they readily admit, they do not formalize any general principles for the domains in question, and the solution remains limited to the scope of the sketch system sKEA. In addition, no justification, in the form of an argument/proof, is provided and verified. We pursue formal models and automated reasoning to solve Bennett problems.

Our Approach and Experiments
We now explain our approach to QMPS via BMCT using a few case studies. Subsequently, we report time-based performance of one of our novel algorithms ($A_B$) in a family ($A_B$) of such for QMPS as required by BMCT.

A Motivating Example
Human adults capable of common-sense reasoning often don’t need to recall exact formulae from physics for everyday mechanical challenges. Our general background, common-sense knowledge usually suffices. Imagine the following scenario:

Your child decides to test you with a simple problem from a Bennett practice test. See Figure 1. Will you try to remember some formulae from physics to solve this? No sane human will do that even if she has excellent memory. Actual tests are timed, and kids have little patience while testing others. So, you only have sufficient time to fall back on your background knowledge, the problem description, and the options. You probably start with eliminating the distractors, one by one, thereby moving toward the most likely key. You know that two cogs of different size cannot rotate with the same speed. What is the relation between size of cog and its speed? Smaller the cog, faster it rotates? Yes! It is indeed cog B which rotates most number of times an hour because it is the smaller one. And, you can explain the answer to your kid with a valid justification!

![Figure 1: A Cog Problem from BMCT with Solution](image)

We believe that an artificial agent too can be engineered to reason as we did in the previous problem. This agent would be well on its way toward capable QMPS in real-life situations.

Formal Mechanical Domain for QMPS
Formalizing background principles in the mechanical domain is an important and early step toward automated reasoning capable of solving qualitative-mechanical problems in general, and BMCT problems in particular. For this, we first use first-order logic (FOL) with a suitable set of relations, functions, and constant symbols. With a straightforward signature (illustrated below), we formalize the relevant principles to build out the formal approach called out (see above) but not pursued in (Klenk et al. 2011).

| Mechanical Domain Signature |
|----------------------------|
| Unary relations = \{Container, Liq, Obj, Swing, Pendulum, Gear, Cog\} |
| Unary functions = \{mass, den, vol, length, radius\} |
| Binary relations = \{Similar, Same, FasterThan\} |
| Binary functions = \{floatsAt, liqLev\} |

Of course, this domain will expand substantially as our AI solves more and more QMPS problems.

Case Studies
Now to two case studies via two actual qualitative-mechanical test problems.

Swing Problem
Figure 2 uses a swing problem to demonstrate how an agent named RealM might plausibly reason about the problem to find a solution, and a cogent supporting argument. A swing is a common example of a pendulum, used frequently to teach pre-college students about the properties and behavior of pendulums. Following a pendulum’s property, the time necessary for a swing to complete 3 cycles of the pendulum’s motion to occur (the period) is directly proportional to the length of the string. A detailed scientific explanation (which by definition exceeds QR) can be found here: https://www.scientificamerican.com/article/bring-science-home-swinging-pendulum.
one cycle is directly proportional to the length of the swing; hence, the smaller the swing length the faster it is. RealM uses the question, context, and options to form its hypothesis and goal and reasons to the correct result, with justification constituted by the reasoning obtained.

RealM reasons about the problem to come up with a solution along with his argument for the solution.

(a) Swing Problem: A Qualitative Mechanical Problem from TEDS Data Dictionary

(b) Agent RealM Reasons Swing Problem.

Figure 2: Case Study I with Swing Problem

Figure 3 shows the complete problem (along with the background principles) represented in FOL using our signature and automatically solved in the HyperSlate® intelligent hypergraphical proof-/argument-assistant system, original kernel of which is conveyed in (Bringsjord et al. 2008).

Figure 3: Swing Problem Solved in HyperSlate®

Liquid Density Problem  Figure 4 shows a QMPS problem from the BMCT’s hydraulics category. By common-sense reasoning, one knows that if two similar containers hold same level of liquids, then volume of the liquids are equal. For equal volume of two different liquids, an object floats higher in the liquid that is denser. Likewise, agent

(a) Liquid Density Problem: A Qualitative Mechanical Problem from Bennett Mechanical Comprehension Test. https://www.iprep.online/courses/bennett-mechanical-comprehension-test-bmct-ii/

(b) Agent RealM Reasons Liquid Density Problem.

Figure 4: Case Study II: Liquid Density Problem

Next, the complete problem (along with the background principles) is represented in FOL using the aforementioned signature and automatically solved in the HyperSlate® proof assistant system. See Figure 5.

Figure 5: Liquid Density Problem Solved in HyperSlate®

Figure 6 shows the problem solved in ShadowProver (Govindarajulu and Bringsjord 2017) using the aforementioned algorithm $A_{B1}$, discussed in more detail in the next subsection.

Liquid Density Problem  Figure 4 shows a QMPS problem from the BMCT’s hydraulics category. By common-sense reasoning, one knows that if two similar containers hold same level of liquids, then volume of the liquids are equal. For equal volume of two different liquids, an object floats higher in the liquid that is denser.

4The concept/relation ‘faster’ here is supposed to mean to the test-taker “greater number of swings/oscillations in a set time period,” and this implies a need, ultimately, for significant NLU science/technology.

5Presentation of complete proof omitted for economy.

6A detailed scientific explanation can be found here: https://www.middleschoolchemistry.com/lessonplans/chapter3/lesson4.

7All automated as well as full manual proofs can be found here: https://github.com/ShreyabBanerjee/QMPS/tree/main/Proofs.

8ShadowProver is a fast automated reasoner available under an open-source license at https://github.com/naveensundarg/prover.
QMPS Core Algorithm

To describe $A_{B1}$, in the family ($\mathcal{A}_B$) of such algorithms for QMPS as required by the format of Bennett/BMCT problems, assume we have access to any first-order theorem prover. While ShadowProver covers higher-order and modal logic, we use it here in “humble” form; denoted by ‘SP.’ For a set of formulae:

- $B$, representing the set of all background principles from each category (c) such as $B_f$ for fluid, $B_p$ for pendulum, $B_g$ for gears, etc.
- $S$, representing the story
- $Q$, representing the question stem
- $\pi_i$, representing each of the options for $i = 1$–$n$ ($n$ can range from 3–5)

let us assign $\Gamma = B_c \cup S \cup Q$, then $\text{SP}(\Gamma, o_i)$ gives us a proof for $\Gamma \vdash o_i$ if $o_i$ is the correct answer/key (denoted by $o_i^*$); otherwise returns FAIL. Here we assume there is only 1 correct option to each question in accordance with BMCT. The following core algorithm for QMPS prints and saves both the correct and incorrect options, along with a proof for the correct option $o_i^*$:

**Input:** Background principles $B$, Category $c$, Story $S$, Question Stem $Q$, Options $o_1$ to $o_n$

**Output:** Stores and prints a proof of $\Gamma \vdash o_i^*$ for the correct option $o_i^*$, otherwise stores and prints $o_i$ as incorrect option

1. Initialization:
   - Analyze Story $S$ and Question Stem $Q$;
   - Select Background Principle $B_c$ from the set of principles $B$ given the Category $c$;
   - $\Gamma = B_c \cup S \cup Q$;
   - $i = 0$;

2. While $i < n$ do
   - $\text{answer} = \text{SP}(\Gamma, o_i)$;
     - If $\text{answer} \neq \text{FAIL}$ then
       - Store $\text{answer}$ and $o_i^*$ as the correct option;
       - $i = i + 1$;
     - While $i < n$ do
       - Store $o_i$ as the incorrect option;
       - $i = i + 1$;
   - Else
     - Store $o_i$ as the incorrect option;
     - $i = i + 1$;

**Algorithm 1:** QMPS Core Algorithm.v1: $A_{B1}$

**Implementation** We implemented algorithm $A_{B1}$ for the liquid density problem presented above, using ShadowProver. The correct option/key accompanied by underlying proof/arguments, is found in 32.76 seconds. See Fig. 6.

**Conclusion, In-Hand & Future Work** We have sought to convey our formal approach to BMCT and thereby — under the hypothesis of Psychometric AI — to QMPS, and progress therein; this complements and extends the work of (Klenk et al. 2011) and will expand as our AI solves more and more problems. We do have in hand nonmonotonic algorithms beyond $A_{B1}$, and will next be assessing and reporting them. Beyond that, hybridization of our logicist algorithms with transformer models has promisingly commenced in our lab.

**Acknowledgments**

We are grateful to the reviewers for their helpful suggestions (not all of which, alas, could under space constraints be heeded). This research is partially enabled by support from ONR and AFOSR (Award # FA9550-17-1-0191).

**References**

Bennett, G. K. 2008. Bennett Mechanical Comprehension Test. Pearson. Retrieved from http://talentlens.pearsonpsychcorp.com.au/files/BMCT_ST_Manual 281:29.

Bringsjord, S., and Schimanski, B. 2003. What is Artificial Intelligence? Psychometric AI as an Answer. In Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI–03), 887–893. San Francisco, CA: Morgan Kaufmann.

Bringsjord, S.; Taylor, J.; Shilliday, A.; Clark, M.; and Arkoudas, K. 2008. Slate: An Argument-Centered Intelligent Assistant to Human Reasoners. In Grasso, F.; Green, N.; Kibble, R.; and Reed, C., eds., Proceedings of the 8th International Workshop on Computational Models of Natural Argument (CMNA 8), 1–10. Patras, Greece: University of Patras.

Davis, E. 2008. Pouring Liquids: A Study in Commonsense Physical Reasoning. Artificial Intelligence 172(12-13):1540–1578.

Forbus, K. 2019. Qualitative Representations: How People Reason and Learn about the Continuous World. Cambridge, MA: MIT Press.

Govindaraju, N., and Bringsjord, S. 2017. On Automating the Doctrine of Double Effect. In Sierra, C., ed., Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI–17), 4722–4730. International Joint Conferences on Artificial Intelligence.

Hayes, P. 1978. The Naive Physics Manifesto. In Mitchie, D., ed., Expert Systems in the Microelectronics Age. Edinburgh, Scotland: Edinburgh University Press. 242–270.

Hayes, P. J. 1985. The Second Naive Physics Manifesto. In Hobbs, J. R., and Moore, B., eds., Formal Theories of the Commonsense World. Norwood, NJ: Ablex. 1–36.

Klenk, M.; Forbus, K.; Tomai, E.; and Kim, H. 2011. Using Analogical Model Formulation with Sketches to Solve Bennett Mechanical Comprehension Test Problems. Journal of Experimental & Theoretical Artificial Intelligence 23(3):299–327.

Weld, D. S., and De Kleer, J. 2013. Readings in Qualitative Reasoning about Physical Systems. Morgan Kaufmann.