האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

SOLVING STUMPERS, CRT AND CRAT: ARE THE ABILITIES RELATED?

By

MAYA BAR-HILLEL, TOM NOAH, AND SHANE FREDERICK

Discussion Paper # 729   (October 2019)

מרכז פדרמן לחקר הרציונליות

THE FEDERMANN CENTER FOR THE STUDY OF RATIONALITY

Solving stumpers, CRT and CRAT: Are the abilities related?

Maya Bar-Hillel*   Tom Noah[†]   Shane Frederick[‡]

Abstract

Bar-Hillel, Noah and Frederick (2018) studied a class of riddles they called stumpers, which have simple, but curiously elusive, solutions. A canonical example is: "Andy is Bobbie's brother, but Bobbie is not Andy's brother. How come?" Though not discussed there, we found that the ability to solve stumpers correlates significantly with performance on items resembling the CRT (Cognitive Reflection Test) but not with performance on items from the CRAT (Compound Remote Associates Test). We report those results here.

Keywords: stumpers, riddles, problem solving, CRT, CRAT, insight problems, cognitive ability, representation

1 Introduction

Bar-Hillel, Noah and Frederick (2018) discussed a class of riddles that challenge respondents to explain a situation which – at first blush – seems impossible or paradoxical. We called these riddles *stumpers*, because respondents often cannot fathom what they are missing, or what alternate representation might permit a solution. Stumpers are a subset of a broader class of insight problems (Gilhooly & Murphy, 2005), called "single step insight problems" by Murray and Byrne (2013).

We liken these stumpers to a play's script, with our mind as the director arranging the scene. Subjects are stumped if the scene they first construct (their "mental model"; see Craik, 1943; Johnson-Laird, 1983) does not contain the solution, and they remain stumped until they are able construct a different scene that can accommodate the script's elements. The four stumpers we used are reproduced in Table 1[1], with solutions and the alternate representations that afford them shown in the Appendix. We extend our earlier paper here, by examining how the ability to solve stumpers correlates with two other types of tasks: the CRT and the CRAT.

Frederick (2005) studied a class of problems that he termed the "Cognitive Reflection Test" (or CRT). For these items, respondents never feel stumped: they typically come up with an answer immediately, offer it without hesitation, and are later surprised to learn it is not correct. The four items we used (see Table 2) are not from the original CRT but share many of the same properties[2]. Except for Mary, three numerical answers, plus an option for "Other", were provided for each item (shown in the Appendix).

Our subjects also answered four items from the Remote Associates Test (or RAT; Mednik, 1962, 1968), in which respondents seek a fourth word that is associated with each of three presented words (see Table 3). We used a more restrictive variant of this test, called the Compound Remote Associates Test or CRAT (Bowden & Jung-Beeman, 2003), As explained in the instructions: "In the next task, you will be presented with three words. You will have 15 seconds to think of a word that forms a word-pair with each of them," and an example was provided.

We included these additional tests because we thought all required some form of creativity, defined by Mednik (1962) as "the forming of ... elements into new combinations which ... are in some way useful" (p. 221). Stumpers typically require subjects to visualize the narrative elements in new ways. As its name implies, the RAT requires respondents to search beyond the immediately available associations until they can find one that all three stem words share (Barr, Pennycook, Stolz & Fuselgang, 2015). Finally, the CRT items generally require respondents to subject their initial intuitive solutions to a subsequent search for potentially disqualifying observations: Mary herself counts as one of her mother's children; the bear lost 20% of its *pre*-hibernation weight, when it weighed *more than* 1000 pounds; the food in a trough will be consumed faster when more animals feed from it, so certainly faster than 6 days; the 15th tallest and 15th shortest individual is the same person – Jerry – so simply adding those numbers will double count him.[3]

---

*Department of Psychology, The Hebrew University of Jerusalem. Email: msmaya@math.huji.ac.il

[†]Department of Psychology, The Hebrew University of Jerusalem.

[‡]Department of Marketing, Yale University School of Management.

[1]The wording has since been improved, but we report the original wording here.

---

[2]Mary was used by Thomson & Oppenheimer (2016), as a type of non-mathematical CRT which they called CRT-2.

[3]CRT problems frequently involve up to 3 cognitive stages: (1) the production of an intuitive response upon exposure to the problem; (2) noticing something that disqualifies that answer; and (3) a search for an alternate solution. In marked contrast, stumpers involve 0 stages (if one remains stumped) or 1 (if the right representation occurs).

TABLE 1: Four stumpers. Solution rates in parentheses.

Accountant N=99 [48%]	An accountant says: 'That attorney is my brother', and that is true – they really do have the same parents. Yet that attorney denies having any brothers — and that is also true! How is that possible?
Speeding Car N=99 [39%]	A big brown cow is lying down in the middle of a country road. The street lights are not on, the moon is not out, and the skies are heavily clouded. A truck is driving towards the cow at full speed, its headlights off. Yet the driver sees the cow from afar easily, and avoids hitting it, without even having to brake hard. How is that possible?
Potato Bags N=95 [38%]	In a Bangladesh market, a small potato bag costs 5 taka, a medium potato bag costs 7 taka, and a large potato bag costs 9 taka. Yet, a single potato in that market costs 10 taka. How is that possible?
Bus Ride N=101 [12%]	Individual bus rides cost one dollar each. A card good for five rides costs five dollars. A first-time passenger boards the bus alone and hands the driver five dollars, without saying a word. Yet the driver immediately realizes, for sure, that the passenger wants the card, rather than a single ride and change. How is that possible?

2 Method

The four Study 2 stumpers were answered by 394 respondents, recruited on Amazon's Mechanical Turk. Their mean age was 38, and 55% were female. Respondents were randomly assigned to one of four experimental groups, defined by the four stumpers in Table 1, and paid a dollar to participate. They answered a multi-screen questionnaire, administered through Qualtrics. The questionnaire began with one of our four stumpers, and ended with all the CRT items and RAT items shown in Tables 2 and 3[4] (full details can be found in Bar-Hillel et al., 2018). Thus, although each stumper was answered by only about 100 respondents, all

---

[4]CRT and CRAT items were presented in two different orders: {Mary, Bear, Jerry, Food} with {AHC, RBS, CFT, WMS} or {Mary, Food, Jerry, Bear} with {RBS, AHC, WMS, CFT}.

TABLE 2: CRT items. Solution rates in parentheses.

Mary [57%]	Mary's mother has four children. She named the youngest three Spring, Summer, and Autumn. What is the name of the oldest child?
Bear [52%]	A bear lost 20% of its weight during the winter hibernation. When it emerged from the hibernation, it weighed 1000 pounds. How much did it weigh before the hibernation?
Food [49%]	A trough of food can feed Anne's flock for 6 days. It can feed Ben's flock for 12 days. How long will it last if both flocks feed from it together?
Jerry [32%]	Jerry's teacher measured all the kids in Jerry's class. Jerry came in 15th tallest as well as 15th shortest. How many kids are there in that class?

TABLE 3: CRAT items. Solution rates in parentheses.

Item	Stem words		
WMS [64%]	Walker	Main	Sweeper
CFT [59%]	Chocolate	Fortune	Tin
RBS [56%]	Room	Blood	Salts
AHC [57%]	Ache	Hunter	Cabbage

394 received the four cognitive reflection items and the four remote associate items. Respondents were allowed to leave any item they wished unanswered.

3 Results

We scored respondents' stumper performance as 0 for "failed", and 1 for "solved". We scored their performance on the two other tasks by the number of items they solved (from 0 to 4). Cronbach's alpha was .49 for the CRT, and .50 for RAT, hence essentially identical.

Figure 1 shows the scores on these other scales when respondents are split by success [or failure] on their stumper. Those who solved their stumper scored significantly higher on our four CRT items, but not on our four-item CRAT.

Table 4, in addition to the size of these effects, shows the associated point-biserial correlation between the scales (r). Solving the stumper predicted solving the CRTs (r(394) =
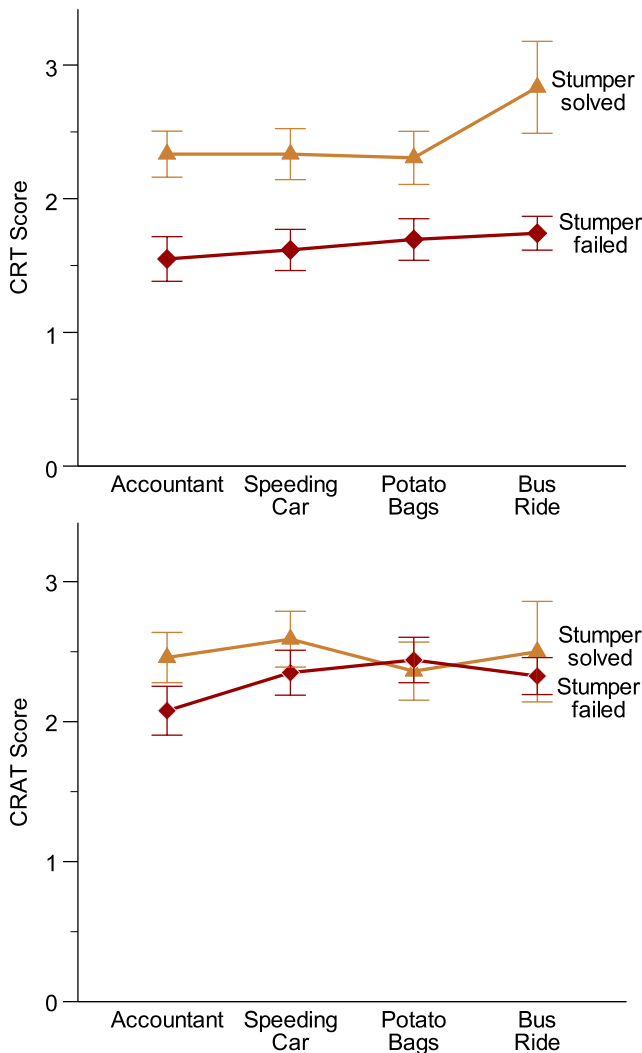
FIGURE 1: CRT and CRAT scores for Ss who solved, or failed, their stumper. Error bars are s.e.'s.

.27, p <.001), but did not predict solving the CRATs (r(394) = .063, p = .210). These two correlations differ significantly (Z = 3.310, p < .01).

The correlation between the CRT and the CRAT, .187, was highly significant.

Stumper success was not significantly correlated with either age or gender (excepting the Accountant stumper, which was solved by 59% of women, but just 36% of men).

4　Discussion

Our interest in stumpers reflects our belief that they might reveal novel psychological principles, but it remains unabashedly exploratory. The present paper focuses on relations between performance on stumpers and two other types of problems — the CRT and the CRAT, which have each

TABLE 4: Relation between stumper solution and two other scales. D is Cohen's D. Percents correct are in brackets.

Stumper	CRT			CRAT		
	Score	D	r	Score	D	r
Accountant						
Solved [48%]	2.33	.67	.32	2.46	.30	.15
Failed [52%]	1.55			2.08		
Speeding Car						
Solved [39%]	2.33	.58	.28	2.59	.19	.09
Failed [61%]	1.62			2.35		
Potato Bags						
Solved [38%]	2.31	.53	.25	2.36	−.06	−.03
Failed [62%]	1.69			2.44		
Bus Ride						
Solved [12%]	2.83	1.00	.29	2.50	.16	.05
Failed [88%]	1.74			2.32		
Overall						
Solved [34%]	2.37	.60	.27	2.48	.13	.06
Failed [66%]	1.65			2.30		

been studied extensively, and correlate with many other psychological variables (see lists in, e.g., Lee, Huggins & Therriault, 2014, for the RAT; Pennycook, Cheyne, Koehler & Fugelsang, 2016, for the CRT). But, aside from expecting the three tasks to correlate positively (as all intellective tasks do), we had no strong predictions, as all seem to involve a similar sort of skill: the ability or disposition to broaden one's search beyond the elements that are initially most accessible.

Accordingly, we remain perplexed why solving stumpers correlates much more strongly with the CRT than with the CRAT, particularly since versions of those two scales often correlate strongly with – and thereby index – general cognitive ability (see, e.g., Frederick, 2005; Chein & Weisberg, 2014; Lee, Huggins & Therriault, 2014). We assume future research will reveal the essential differences among these related types of problems, and hope stumpers will join the ranks of RAT, CRT, insight problems, and other reasoning tasks as a tool for studying mental processes.

References

Bar-Hillel, M., Noah, T., & Frederick, S. (2018). Learning psychology from riddles: The case of stumpers. *Judgment & Decision Making*, *13*(1), 112–122.

Barr, N., Pennycook, G., Stolz, J.A., & Fugelsang, J.A. (2015). Reasoned connections: A dual-process perspec-

tive on creative thought. *Thinking & Reasoning*, *21*(1), 61–75.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associates problems. *Behavior Research Methods, Instruments & Computers, 35*, 634–639

Chein, J. M., & Weisberg, R. W. (2014). Working memory and insight in verbal problems: Analysis of compound remote associates. *Memory & Cognition*, *42*(1), 67–83.

Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25–42.

Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning*, *11*(3), 279–302.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.

Lee, C. S., Huggins, A. C., & Therriault, D. J. (2014). A measure of creativity or intelligence? Examining internal and external structure validity evidence of the Remote Associates Test. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(4), 446–460.

Mednik, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*(3), 220.

Mednik, S. A. (1968). The remote associates test. *The Journal of Creative Behavior*, *2*(3), 213–214.

Murray, M.A. & Byrne, M.J. (2013). Cognitive change in insight problem solving: Initial model errors and counterexamples. *Journal of Cognitive Psychology, 25*(2), 210–219.

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, *48*(1), 341–348.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Appendix: Solutions

Dominant and alternate representations, with solutions, for our stumpers.

Stumper	Dominant representation (blocks solution)	Alternate representation (yields solution)
Accountant	Male accountant	Female accountant
Speeding Car	Nighttime	Daytime
Potato Bags	Full potato bags	Empty potato bags
Bus Ride	Payment with one $5 bill	Payment with five $1 bills

CRT solutions.

Mary	Mary
Bear	800 lbs; 1200 lbs; 1250 lbs; Other; (Correct: 1250)
Food	4 days; 9 days; 18 days; Other; (Correct: 4 days)
Jerry	29 kids; 30 days; 31 days; Other; (Correct: 29)

CRAT solutions:
WMS: Street; CFT: Cookie; RBS: Bath; AHC: Head

HEBREW UNIVERSITY OF JERUSALEM
מרכז פדרמן לחקר הרציונליות
THE FEDERMANN CENTER FOR THE STUDY OF RATIONALITY

<u>List of Recent Discussion Papers</u>

These and earlier papers can be found on our website: www.ratio.huji.ac.il

694. **Elchanan Ben-Porath, Eddie Dekel, Barton L. Lipman,** Disclosure and Choice (February 2016)

695. **Aviv Keren**, The Logic of Love (March 2016)

696. **David Heyd, Uriel Procaccia, Uzi Segal,** The Thorny Quest for a Rational Constitution (March 2016)

697. **Uriel Procaccia, Eyal Winter,** Corporate Crime and Plea Bargains (March 2016)

698. **Uriel Procaccia, Eyal Winter**, Corporate Bill of Rights (March 2016)

699. **Maya Bar-Hillel**, Reply to Rodway, Schepman & Thoma (2016) (April 2016)

700. **Ilan Nehama**, Analyzing Games with Ambiguous Player Types Using the MINthenMAX Decision Model (August 2016)

701. **Hanan Shteingart, Yonatan Loewenstein**, Heterogeneous Suppression of Sequential Effects in Random Sequence Generation, but not in Operant Learning (september 2016)

702. **Ohad Navon**, Evolutionarily Stable Strategies of Random Games and the Facets of Random Polytopes (september 2016)

703. **Bezalel Peleg, Shmuel Zamir** , Sequential aggregation judgments: Logical derivation of relevance relation (september 2016)

704. **Sophie bade , yannai a. Gonczarowski**, Gibbard-Satterthwaite Success Stories And Obvious Strategyproofness (October 2016)

705. **Shira Cohen-Zimerman, Ran R. Hassin,** Implicit Motivation Makes the Brain Grow Younger: Improving Executive Functions of Older Adults (January 2017)

706. **Abraham Neyman , Elon Kohlberg**, Cooperative Strategic Games (February 2017)

707. **Ohad Dan, Drorith Hochner- Celnikier, Amy Solnica, Yonatan Loewenstein**, Association Of Catastrophic Neonatal Outcomes With Increased Rate Of Subsequent Cesarean Deliveries (April 2017)

708. **Bezalel Peleg, Shmuel Zamir,** Sequential Aggregation Of Judgments (May 2017)

709. **Johannes Müller-Trede, Shoham Choshen-Hillel, Meir Barneron and Ilan Yaniv,** The Wisdom of Crowds in Matters of Taste (May 2017)

710. **Elon Kohlberg and Abraham Neyman,** Games of Threats (September 2017)

711. **Sergiu Hart,** Repeat Voting: Two-Vote May Lead More People To Vote (October 2017)

712. **Sergiu Hart and Philip J. Reny,** The Better Half of Selling Separately (December 2017)

713. **Shevy Waner, Uzi Motro, Yael Lubin, and Ally R. Harari,** Male mate choice in a sexually cannibalistic widow spider (February 2018)

714. **Maya Bar-Hillel, Tom Noah, and Shane Frederick,** Learning psychology from riddles: The case of stumpers (February 2018)

715. **Bezalel Peleg and Ron Holzman,** Representations of Political Power Structures by Strategically Stable Game Forms: A Survey (February 2018)

716. **Kristoffer Arnsfelt Hansen, Rasmus Ibsen-Jensen, and Abraham Neyman,** The Big Match with a Clock and a Bit of Memory (February 2018)

717. **Bezalel Peleg and Hans Peters,** Self-implementation of social choice correspondences in strong equilibrium (April 2018)

718. **Abraham Neyman,** Additive valuations of streams of payoffs that obey the time-value of money principle: characterization and robust optimization (April 2018)

719. **Bezalel Peleg and Shmuel Zamir,** Judgements aggregation by a sequential majority procedure (June 2018)

720. **Lotem Elber-Dorozko and Yonatan Loewenstein,** Striatial Action-Value Neurons Reconsidered (July 2018)

721. **Orli Bobek, Adiv Gal, David Saltz, and Uzi Motro,** Effect of Nest-Site Microclimatic Conditions on Nesting Success in the Lesser Kestrel (Falco Naumanni) (July 2018)

722. **Barry O'Neill,** Two-Party Agreements as Circular Sets (September 2018)

723. **Adiv Gal, David Saltz, and Uzi Motro,** Effect of Supplemental Feeding on Nesting Success in the Lesser Kestrel (Falco Naumanni) (September 2018)

724. **Maya Bar-Hillel,** The unbearable lightness of self-induced mind corruption (November 2018)

725. **Tomer Siedner,** Optimal pricing by a risk-averse seller (May 2019)

726. **Maya Bar-Hillel and Cass R. Sunstein,** Baffling bathrooms: On navigability and choice architecture (June 2019)

727. **Maya Bar-Hillel and Jacob Lavee,** Lay attitudes toward involuntary organ procurement from death-row prisoners: no, but (June 2019)

728. **Maya Bar-Hillel,** Why didn't I see it earlier? (July 2019)

729. **Maya Bar-Hillel, Tom Noah, and Shane Frederick,** Solving stumpers, CRT and CRAT: Are the abilities related? (October 2019)