Cognition 214 (2021) 104288

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Brief article

Still no solution to non-verbal measures of analogical reasoning: Reply to Walker and Gopnik (2017)

G.C. Glorioso, S.L. Kuznar, M. Pavlic, D.J. Povinelli *

The Biology Department, University of Louisiana, Lafayette, LA 70504, United States

ARTICLE INFO

Keywords:
Relational match-to-sample
Analogical reasoning
Nonverbal cognition
Animal cognition

ABSTRACT

Walker and Gopnik (2017) suggest they have solved a longstanding problem in comparative and developmental psychology: How to provide an unambiguous measure of analogical reasoning in nonverbal subjects. We argue that this test, much like many others that purport to measure analogical reasoning in nonverbal subjects, does not distinguish between the two competing accounts of successful performance: the use of perceptual variance among stimuli to support higher-order concepts like “same” and “different” versus use of perceptual variance alone.

In what is described as a “proof of concept”, Walker and Gopnik (2017) suggest they have solved a longstanding problem in comparative and developmental psychology: How to provide an unambiguous measure of analogical reasoning in nonverbal subjects. As Walker and Gopnik (2017) note, analogical reasoning is widely recognized as a foundational aspect of the most distinctive aspects of cognition (see Gentner, Holyoak, & Kokinov, 2001; Penn, Holyoak, & Povinelli, 2008; Premack, 2010; Rumelhart & Abrahamson, 1973; Sternberg, 1977). Indeed, Robert Oppenheimer (1956) underscored the importance of analogical reasoning in scientific progress in his address to the 63rd annual meeting of the American Psychological Association. Thus, developing unambiguous measures of analogical reasoning is of vital importance in understanding its early evolutionary and developmental expression.

Penn et al. (2008) have argued that while many organisms are capable of first-order, perceptually-based relational reasoning, only humans are cable of higher-order, role-based analogical reasoning. Higher-order concepts (such as same/different) may be one of the foundations upon which there are marked functional differences between humans and other species (see Penn et al., 2008). One of the most widely used procedures to attempt to detect the presence of higher-order analogical reasoning in human children and animals has been the *relational* match-to-sample (RMTS) test. In the simplest RMTS test, subjects are initially presented a pair of objects that instantiate a given higher-order relation, such as *same* (AA) or *different* (BC). The subjects are then presented two new object pairs, one of which instantiates the targeted relation. So, on *same* trials, BB or CD follows the initial AA stimulus. Subjects are rewarded if they select the “conceptually” matching pair (i.e., BB). In studies with nonhuman animals, subjects must typically be trained (using differential reinforcement across hundreds or thousands of trials) to select the correct pair (see Flemming, Thompson, & Fagot, 2013). On test trials, new object pairs are presented that are once again perceptually identical or perceptually different. Some researchers have argued that a subject’s ability to select the object pair that instantiates the previously rewarded relationship is evidence that they are using the higher-order relational concept (i.e., *same* or *different*). Variants of the RMTS test have been administered to a variety of organisms including human infants (Hochmann et al., 2017), chimpanzees and rhesus monkeys (Flemming, Beran, Thompson, Kleider, & Washburn, 2008), budgerigars (Manabe, Kawashima, & Staddon, 1995), California sea lions (Schusterman & Kastak, 1993), ducklings (Martinho & Kacelnik, 2016) and even crows (Smirnova, Zorina, Obozova, & Wasserman, 2015).

Unfortunately, as Walker and Gopnik (2017) point out, the results of such tests share an inherent ambiguity: Correct performances can result from either a conceptual understanding of *same* (or *different*) or by tracking the *perceptual entropy* in the rewarded relation. Perceptual entropy has been defined as the amount of inter-item variability in an array (Fagot, Wasserman, & Young, 2001), which is commonly calculated by Shannon and Weaver’s (1949) measure of (informational) entropy:

$$H' = -\sum_{i=1}^{R} p_i \log p_i$$

Applied to an identical two-item array (AA), this equation yields an

---

* Corresponding author.
E-mail addresses: c00304772@louisiana.edu, gabrielleglorioso@yahoo.com (G.C. Glorioso), povinelli@louisiana.edu (D.J. Povinelli).

https://doi.org/10.1016/j.cognition.2020.104288
Received 26 March 2019; Received in revised form 20 February 2020; Accepted 30 March 2020

entropy value of zero (i.e., no perceptual difference between items in the arrays). An array of two different items (BC), in contrast, would yield a value greater than one.[1] Mathematically, for perceptual entropy to exist, there must be at least two items in an array. Thus, the purported *same/different* relation that is at stake in the RMTS test can be understood as "a continuous, analog estimate of the degree of perceptual variability between the elements in each display" (Penn et al., 2008). Thus, the ambiguity inherent in any RMTS test is that a subject can learn to match the perceptual entropy of the items rather than matching the higher-order, role-based analogical relation of *same/different*. For example, if subjects are initially rewarded for selecting a zero-entropy array, they will be correct if they select a novel array that instantiates zero-entropy (and mutatis mutandis when subjects are rewarded to pick sample arrays that instantiate entropy values greater than zero). A fact often downplayed in the literature is that even those organisms that do possess higher-order relational concepts such as *same/different* (i.e., adult humans) must also (by default) possess the ability to detect perceptual entropy. Otherwise, there would be no physiological basis to assess the within-item differences that can then be interpreted as the higher-order *same/different* relation in question. Detecting and using perceptual entropy, then, does not require the use of higher-order relations such as *same/different*. Detecting higher-order relations, however, does require detecting perceptual entropy. This asymmetry may leave no unique causal role for high concepts in traditional RMTS tests.

Walker and Gopnik (2017) developed a new variant of the RMTS test and argue that it can distinguish between a perceptual strategy versus a higher-order relational strategy. In their experimental design, human toddlers (18–30 months) were presented with blocks of various shapes. When the blocks were placed on top of a music box, the box either played enjoyable music or remained silent. The toddlers were randomly assigned to one of two groups, and each group went through an observational learning phase and a test phase.

The first group was labeled the "unfused-object group". In the observational learning phase, toddlers in this group were shown that when two separate but identical blocks (AA) were placed on the box it played music (n = 4 trials), whereas when two separate but different blocks (BC) were placed on the box, it remained silent (n = 4 trials). New pairs of blocks (instantiating the respective relations) were used on each trial. The second group was labeled the "fused-object group". In the observational learning phase for this group, two blocks were glued together and presented as a single object. In this case, the single object that was composed of two identical blocks (i.e., a symmetrical block) made the box play music (n = 4 trials), whereas a single object composed of two different blocks (i.e., an asymmetrical block) did not (n = 4 trials). New (fused) blocks were used on each trial.

In the test phase, new blocks were used for both groups. The toddlers were presented with two trays, one of which contained the "correct" choice (the blocks or block instantiating the relation activated the music box), and another tray contained the "incorrect" choice. The results were straightforward. Subjects in the unfused-object group selected the correct pair at above chance levels. In contrast, subjects in the fused-object group performed at chance levels.

Walker and Gopnik (2017) state that using a *same/different* strategy would lead to success in the unfused object condition, but failure in the fused object condition. On the other hand, a strategy based on "perceptual variation" ought to result in success in both conditions. In their words: "If children are indeed relying upon a low-level perceptual heuristic, they should select the lower entropy pair consistently across both conditions. On the other hand, if children learn the abstract relation 'same,' they should privilege this test pair only in the unfused/relational condition, where these is a relation to learn" (Walker & Gopnik, 2017, p. 24). Because the authors obtained the latter results, they concluded that children were using a higher-order *same/different* concept to solve the unfused condition.

Walker and Gopnik's (2017) logic derives from their idea that (in theory) single objects possess differing amounts of "variance among the features … (i.e., colors, edges, angles)" (p. 26). Although this is not stated explicitly, the logic would appear to be that if the toddlers rely upon "low-level perceptual heuristics", they would initially calculate a freestanding value for a given object (symmetrical or asymmetrical). Then, *between trials* they would compare that value to the next object. No formal specification was offered for how such within-object perceptual variation is assessed in the first place, but the authors do cite a study with pigeons conducted by Young, Wasserman, Hilfers, and Dalrymple (1999). Unfortunately, the pigeons were presented with one object at a time and requested to select *same* or *different* based upon previous presentations, thus requiring the pigeons to detect standard entropy (see above) between objects separated temporally, rather than spatially. Nonetheless, if we assume that a theory of within-object perceptual variation could be defined for the individual objects in the fused object condition, do their results logically bolster the higher-order interpretation of the results of the *unfused* condition? We do not think so.

First, *symmetry/asymmetry* can be considered as yet another type of higher-order relation. But any RMTS-like task that attempted to instantiate *symmetry/asymmetry* would fall prey to the same difficulties of interpretation as RMTS *same/different* tests. This highlights the following question: Why did the toddlers *not* use the higher-order *symmetry/asymmetry* relation to solve the fused condition, let alone the "low-level perceptual heuristics" suggested by Walker and Gopnik (2017)? The difficulty in answering this question arises because even if toddlers did in fact possess higher-order concepts such as *symmetry/asymmetry* or *same/different*, perceptual variance between objects would still need to be calculated (whether within trials or between trials or both). This claim does not feel controversial to us. In the case of visual stimuli, for example, photons bouncing off objects must strike the retina and then be processed sufficiently before any activation/recruiting/mapping of purported *same/different* concepts.

This speaks to the interpretive limit of the current paradigm: the nature of the "perceptual variation/entropy" is not the same in the two conditions. Consider an unfused trial involving *same* pairs. Here, subjects could detect (1) Walker and Gopnik's low *within-object* variability for each of two objects (considered separately), and (2) an additional, low (zero) entropy *between* the two objects. The same analysis applies, mutatis mutandis, for different pairs. Subjects then must relate this information to the relevant contingency (music activation) across trials.

In contrast, consider a fused trial involving an *asymmetrical* object. In this case, the only perceptual variability available to subjects is the within-object variability, and in this case, *for a single object* (only one object is present). Given this experimental design, no additional information (e.g., from another asymmetrical object) is available. So, unlike trials involving *same* pairs, subjects must assess entropy information *across trials* to map the contingency of music activation. This analysis reveals that even without a specific theory of how individual objects are perceived, the fused and unfused conditions are not matched for the amount and type of perceptual variability available in the initial displays. Worse yet, the toddlers succeeded in the condition (unfused) that contained greater, and mutually reinforcing, sources of information

---

[1] One important version of the RMTS test involves two larger arrays involving up to 16 different visual icons which are either all the same or all different. Fagot et al. (2001) compared two adult baboons and two adult human subjects with an RMTS task using such arrays. Both baboons and humans learned to pass the RMTS test and successfully generalized to novel sets of stimuli. However, when the number of items in the arrays was systematically reduced from 16 to 2, the performances fell to chance on *different* arrays but remained above chance on *same* arrays. Other variants have attempted to use different degrees of *same* or *different* within arrays containing multiple objects (i.e., AAAA, AAAB, AABC, ABCD; Flemming et al., 2013). Importantly, Wasserman and colleagues have long suggested that signatures of entropy can be detected in such tasks (review by Wasserman & Young, 2010; see also Hochmann et al., 2017). Although our interpretation of these signatures differs, this issue is not germane to the issues at stake here, so it is not considered further.

related to perceptual variability. We stress that our analysis says nothing about whether such perceptual variation activates concepts of *same/different* (or even *symmetry/asymmetry*) in these toddlers. Rather, it illustrates why we believe this between-condition comparison cannot say anything about it either.

Is it possible the toddlers in the fused-object group were intended to represent the fused objects as two separate objects, and thus create an implicit relation between the two? This seems unlikely. If the toddlers represented the fused objects as being composed of two objects, then the fused condition would be identical to the unfused condition and an entropy value could be calculated for each stimulus. In that case, the entropy and *same/different* accounts would generate the (incorrect) prediction of success in both conditions.

Walker and Gopnik's (2017) experimental procedure does not escape the inherent ambiguity of RMTS test as evidence of analogical reasoning. Like all existing same/different and RMTS tasks, their new procedure is solvable by the detection of perceptual variability between or among stimuli. The fact that some methods of presenting this variability make it harder to detect and process does not change this conclusion. This underscores the pressing need to develop alternative procedures for measuring the expression of higher-order analogical reasoning in human infants and other species.

Credit author statement

Gabrielle Glorioso: Investigation; Writing – Original Draft; Writing – Reviewing & Editing. Shannon Kuznar: Investigation; Writing – Reviewing & Editing. Mateja Pavlic: Investigation; Writing – Reviewing & Editing. Daniel Povinelli: Investigation; Writing – Reviewing & Editing; Supervision; Project administration.

References

Fagot, J., Wasserman, E., & Young, M. E. (2001). Discriminating the relation between relations: The role of entropy in abstract conceptualization by baboons (*Papio papio*) and humans (*Homo sapiens*). *Journal of Experimental Psychology Animal Behavior Processes, 27*, 316–328.

Flemming, T. M., Beran, M. J., Thompson, R. K. R., Kleider, H. M., & Washburn, D. A. (2008). What meaning means for same and different: Analogical reasoning in humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), and rhesus monkeys (*Macaca mulatta*). *Journal of Comparative Psychology, 122*, 176–185.

Flemming, T. M., Thompson, R. K. R., & Fagot, J. (2013). Baboons, like humans, solve analogy by categorical abstraction of relations. *Animal Cognition, 16*, 519–524.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (Eds.). (2001). *The analogical mind: Perspectives from cognitive science*. Cambridge, MA: MIT press.

Hochmann, J. R., Tuerk, A. S., Sanborn, S., Zhu, R., Long, R., Dempster, M., & Carey, S. (2017). Children's representation of abstract relations in relational/array match-to-sample tasks. *Cognitive Psychology, 99*, 17–43.

Manabe, K., Kawashima, T., & Staddon, J. E. R. (1995). Differential vocalization in budgerigars: Towards an experimental analysis of naming. *Journal of the Experimental Analysis of Behavior, 63*, 111–126.

Martinho, A., & Kacelnik, A. (2016). Ducklings imprint on the relational concept of "same or different". *Science, 353*, 286–288.

Oppenheimer, R. (1956). Analogy in science. *American Psychologist, 11*, 127–135.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 109–178.

Premack, D. (2010). Why humans are unique: Three theories. *Perspectives on Psychological Science, 5*, 22–32.

Rumelhart, D. E., & Abrahamson, A. A. (1973). A model for analogical reasoning. *Cognitive Psychology, 5*, 1–28.

Schusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record, 43*, 823–849.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology, 25*, 256–260.

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review, 84*, 353–378.

Walker, C. M., & Gopnik, A. (2017). Discriminating relational and perceptual judgments: Evidence from human toddlers. *Cognition, 166*, 23–27.

Wasserman, E. A., & Young, M. E. (2010). Same–different discrimination: The keel and backbone of thought and reasoning. *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 3–32.

Young, M. E., Wasserman, E. A., Hilfers, M. A., & Dalrymple, R. (1999). The pigeon's variability discrimination with lists of successively presented visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 475–490.