AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2022 American Psychological Association
ISSN: 0096-3445

2022, Vol. 151, No. 1, 224–245
https://doi.org/10.1037/xge0001068

The Importance of Inference in Relational Reasoning: Relational Matching as a Case Study

Ivan G. Kroupin and Susan E. Carey
Department of Psychology, Harvard University

Nonhuman animals and preschoolers struggle with Relational-Match-to-Sample (RMTS), a classic test of the capacity for second-order relational, analogical, and reasoning. These failures are often explained by limitations in representational or computational capacities. Drawing on recent evidence for robust spontaneous RMTS success (i.e., without error-feedback) in crows and parrots after minimal second-order training, we present five experiments with human adults consistent with the possibility that population differences sometimes instead derive from differences in inductive biases alone. Experiment 1 confirms human adults have the capacities and requisite representations to succeed spontaneously on RMTS. Experiments 2–5 utilize a modified RMTS task in which adults make relational matches only about half of the time. Experiment 3 tests whether eight trials of various MTS tasks, *nonsecond-order* training featured in the aforementioned comparative studies, can increase spontaneous *second-order* relational responding in human adults. Two of the MTS tasks (Number, Size MTS) do so, demonstrating that MTS training can, in fact, increase relational responding by changing inductive biases alone. The other MTS tasks (Identity, Color MTS) do *not* do so, evidence that the facilitating effect is not a result of matching involved in MTS per se. Experiments 4 and 5 test one hypothesized mechanism by which specifically Number/Size MTS tasks may have led to increased relational responding, that is, by inhibiting preexisting biases to match on shape and/or color, making relational matches relatively more likely. We close by discussing the importance of research into inductive biases to the project of understanding relational reasoning.

Keywords: relational reasoning, comparative cognition, abstract relations, inductive reasoning, relational match to sample

> All the world's a stage, And all the men and women merely players.
>
> —(Shakespeare, 1963, 2.7.1037)

Shakespeare's comparison between actors on a stage and people in the world, and our ability to understand it, reflects the capacity to reason about and compare the relations holding within different sets of individual entities. This kind of relational reasoning is a cornerstone of human cognition, ubiquitous within ordinary language, and underlies artistic metaphor and scientific analogy. Therefore, characterizing the computational underpinnings of this capacity, as well as accounting for its origins (both over evolution and ontogenesis), are central projects within cognitive science (e.g., Holyoak & Thagard, 1995; Halford et al., 2010; Kotovsky & Gentner, 1996; Premack, 1983).

On "Relational Reasoning:" Dramatic Differences Among Populations Over Phylogeny and Ontogeny

The fact that reasoning involves relations at some level of description is not sufficient to qualify it as relational reasoning in the sense we are interested in here. All complex animals represent relations—such as dominance relations, spatial relations, relations between individuals within representations of events. Analogies and metaphors, however, require what the animal cognition literature calls "second-order" relational reasoning (Wasserman et al., 2017); that is, recognizing that two pairs of individuals stand in the same relation with each other, abstracting away from the individuals in each pair (e.g., bird:nest: bee:?) Relational-Match-to-Sample (RMTS), developed by Premack (1983), is the paradigmatic task used to probe whether young children and nonhuman animals are capable of such reasoning, asking whether they can learn to match pairs of distinct entities on the basis of sharing the relations same or different (see Figure 1: A A goes with B C or D D *or* X Y goes with Z Z or P Q). Clearly, solving RMTS requires

---

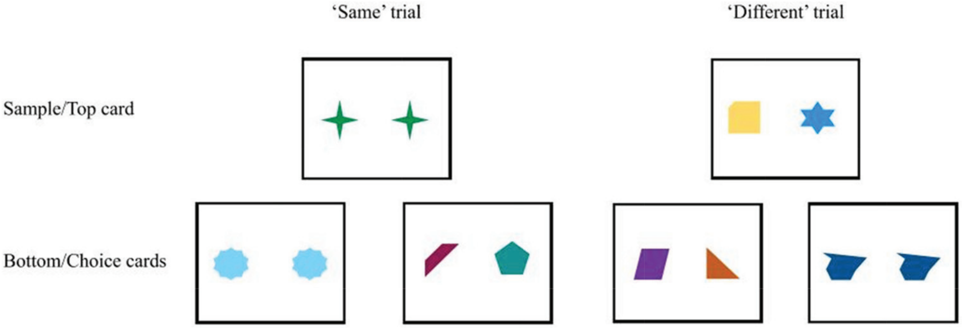
Ivan G. Kroupin https://orcid.org/0000-0003-3673-9918

All the current work has been presented at multiple conferences and workshops including the Budapest Central European University Conference for Cognitive Development, the Analogy 2017 conference, the annual meeting of the Cognitive Science Society, the biannual meeting of the Society for Research in Child Development and the annual meeting of the Human Evolution and Behavior Society. We thank the James S. McDonnell Foundation (Grant 2200204418) and Harvard University for financial support of this work. Original data and materials are available at Open Science Framework (https://osf.io/wt6ne/?view_only=5e9fced2ec4540e8a1ffc692f7724e98).

Correspondence concerning this article should be addressed to Ivan G. Kroupin, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, United States. Email: ikroupin@g.harvard.edu

Figure 1

Examples of Two Relational-Match-to-Sample (RMTS) Trials



Note. See the online article for the color version of this figure.

second-order relational reasoning, as it requires computing relations between relations—that is, whether two pairs both instantiate the relation same or both instantiate the relation different, despite not sharing any individual objects in common.

While the matches involved in Premack's RMTS may appear simple (and are, in fact, extremely easy for human adults in a U.S. sample, see Experiment 1), success on the task has proven to be a tremendous challenge for nonhuman animals (see Wasserman et al., 2017 for a review) and children under the age of about five (see Premack, 1983; Hochmann et al., 2017; Kroupin & Carey, 2020, for evidence of RMTS failures at 4 years).[1] None of these populations succeed spontaneously (with no training or error feedback) on RMTS, and most animals fail even with tens of thousands of reinforced trials of training on the task (though see below for important exceptions), while children aged four and under fail even after eight trials with correct/incorrect feedback (Hochmann et al., 2017). In contrast, both animals as simple as honeybees (Giurfa et al., 2001) as well as infants as young as 10 months (Hochmann et al., 2016) relatively quickly learn MTS tasks (i.e., A goes with B or A). That is, these populations understand matching tasks in general, but many more complex animals and much older children fail to match on the basis of relations, that is, engage in second-order relational reasoning (Premack, 1983; see Wasserman et al., 2017 for a review).

Distinguishing Match Computations From Mental Representations of the Relation Same

Prima facie both RMTS and MTS are relational tasks since both involve the relation same—identifying two entities as the same on some feature in MTS, and recognizing that two distinct pairs of objects instantiate the same internal relations in RMTS. Wasserman et al. (2017) refer to MTS as a "first-order relational reasoning" task, presupposing that there is some representation, some symbol with the content *same*, involved in solving the task.

However, many researchers have noted that MTS and non-MTS (nMTS, i.e., A goes with B not A) can be solved without representing any rule that requires a mental symbol for the relations same or different (Hochmann et al., 2016; Premack, 1983; Zentall et al., 2018). Many cognitive processes involve match computations, including all acts of recognition and categorization. When a chimpanzee or a human recognizes her baby, she matches mental representation of the currently perceived entity with a stored representation. This involves a match computation, a computation of sameness, but does not require a mental symbol for sameness. The only mental symbols required are a stored representation of her baby's features and a representation of the currently attended-to entity, each of which enters into some feature comparison process with an appropriate threshold for recognition.

Success on MTS could be achieved by establishing a program: (store x, seek x), where x is a variable to be filled by representations of arbitrary samples. The only symbols in this procedure are *store*, *choose*, and *x* (a working memory representation of a sample). Analogously, nMTS can be achieved by (*store x*, *avoid x*). Furthermore, Hochmann et al. (2016) and Zentall et al. (2018) provide evidence consistent with these procedures actually underlying the success of 10-month-old infants' and pigeons', respectively, on both MTS and nMTS. In contrast, second-order relational reasoning, including RMTS, *does* require a symbol for the relation same: The variable *x* must be filled by mental representations with the content *same* and/or *different.* We designate such mental symbols here *same*, to portray that the meaning of the mental symbol is same, or Ω, to portray the hypothesis that there is a single arbitrary symbol with this meaning implemented in the mind in some other, unknown way. For humans over age three (at least in a U.S. population, Hochmann et al., 2020), one abstract representation of sameness is literally *same*, that is, a mental representation of the word "same" in the mental lexicon. Another example is the mental representation of a heart-shaped figure that chimps were taught to recognize as mapping to pairs of identical objects (Thompson et al., 1997).

In summary, two facts are apparent from the (R)MTS literature: (1) There is an in-principle difference between MTS and RMTS such that, unlike matching, say, red to red (as in MTS), matching same to same and different to different (as in RMTS) *must* involve the kind of second-order relational reasoning which underpins the human capacity for analogy and metaphor. (2) There is an empirical difference between the difficulty of MTS and RMTS: The former is solved by animals as simple as bees and 10-month-old

---

[1] Other research has found success at earlier ages, but only using simplified RMTS paradigms (e.g. Christie & Gentner, 2014; Walker & Gopnik, 2014). See Kroupin and Carey (2020) for a discussion of RMTS variants.

humans, the latter is *at best* a serious challenge even for primates and human children as old as five (Hochmann et al., 2017; Penn et al., 2008; Premack, 1983; Thompson & Oden, 1995; Wasserman et al., 2017). The outstanding question, then, is what these facts imply for the emergence of second-order relational reasoning capacities in phylogeny and ontogeny.

Plan of the Current Article

The current article begins by reviewing four accounts of why RMTS is so difficult for species other than humans and young children and brings to bear results from a representative sample of training studies in deciding which accounts are tenable in light of existing evidence. To preview, we argue that all four accounts are likely to be true for some population differences on some tasks. With respect to RMTS in particular, we argue (a) that successful training studies make it implausible that *all* population differences are a result of fundamental discontinuities between the cognitive capacities of those who succeed and fail on RMTS and (b) that it is possible—but not yet conclusively demonstrated—that some population differences are *entirely* a result of differences in inductive biases alone. We then present studies that, **first**, conclusively demonstrate that certain MTS training tasks drawn from the animal literature can increase relational responding by changing inductive biases *alone*, **second**, provide further evidence (in line with Hochmann et al., 2016; Zentall et al., 2018) that MTS tasks do not, in fact, involve the same representations as RMTS and, **third**, begin to explore the *mechanisms by which* MTS training tasks can change inductive biases relevant to second-order relational reasoning despite not involving the same representations.

Accounting for the Difficulty of RMTS

Researchers have offered four broad classes of explanations for the failures of some populations on RMTS. The first two explain failures in terms of *capacity limitations* of those populations who fail. The second two describe population differences—at least in some cases—as differences in *learning histories* and not differences in representational or computational capacities.

Capacity Accounts

Account 1: Differences in Representational Capacity. The first account, championed by Penn et al. (2008), posits population differences in representational capacity. They propose that "only human animals possess the representational processes necessary for systematically reinterpreting first-order perceptual relations in terms of higher-order, role-governed relational structures" (p. 110), a capacity they argue emerges only in humans after a certain age. As such, they propose that nonhuman animals and young children fail RMTS because they *cannot* generate abstract relational representations—in this case of sameness or difference—"which are (1) independent of any particular source of stimulus control, and (2) available to serve in a variety of further higher-order inferences in a systematic fashion" (p. 112). That is, Penn et al. (2008) propose that these populations lack the processes which would allow them to generate such representations *altogether*.

Crucially, Penn et al. (2008) are not claiming an absence of *any* relational processing from the cognitive systems of nonhuman animals and young children: As discussed above, match computations in the service of recognition are supported by even very simple cognitive systems. However, their claim presupposes that feature matching in acts of recognition does not involve any relational representations—in line with the Hochmann et al. Zentall et al. proposals that MTS and nMTS require only match computations while RMTS requires abstract mental symbols for sameness and difference that support relational reasoning.

In summary, Penn et al.'s hypothesis is that nonhuman animals lack the capacity to form mental representations such as *same* or Ω. We have no doubt that this hypothesis is true for some animals—placozoa and purple vase sea sponges, for instance. The question, then, is whether there is reason to assume failures of complex animals on RMTS—for example, vertebrates, including nonhuman primates and human children—similarly reflect representational capacity limitations, as Penn et al. hypothesize.

Account 2: Differences in Computational Capacity. A second account of population differences in performance on relational reasoning tasks concerns *computational capacity*—for instance limits on executive functions such as working memory (required to hold both relata in mind during comparison, e.g., Halford, 1993) or inhibition (e.g., of attention to nonrelational properties of stimuli, e.g., Richland et al., 2006). For instance, an individual may be able to *represent* sameness and difference in an abstract, human-adult way, but lack the working memory capacity required to hold the representations of these relations in mind and flexibly compare among them. Such computational capacity differences among populations are well-attested; for example, differences in executive function are extensively documented across both species (Maclean et al., 2014) and over ontogenesis (Diamond, 2013), and in principle could make it impossible for some animal species to succeed on RMTS.

Learning History Accounts

Account 3: Developing Specific Relational Representations. Even if an individual is perfectly *capable* of generating and manipulating abstract relational representations of the kind described by Penn et al. (2008) this does not necessarily mean that they have had the learning experience required to *actually generate* representations like *same* or Ω.[2] There are a number of learning mechanisms evidenced in the literature by which new relational representations can be generated, such as progressive alignment (Kotovsky & Gentner, 1996); Quinian bootstrapping (Carey, 2009); and conceptual combination of existent relational primitives (e.g., Fodor, 1975). If abstract representations of sameness and difference of the kind required for RMTS success are not innate, an individual may fail the task—despite being *capable* of constructing such representations—because they have not gone through a process which would *generate* them. What differentiates

---

[2] In practice, the format of available representation (Accounts 1 and 3) and requisite computational capacities (Account 2) are interrelated as the format of representation used by the individual affects the computational demands of the relational reasoning process. For instance, if one has abstract summary symbols for sameness and difference, e.g., the words "same" and "different," one can transforms a six-item RMTS task (e.g., match A A to either B B or C D) into a three-item MTS task to (match "same" to either "same" or "different"), lessening working memory load (see Halford, 1993; Thompson et al., 1997). Nevertheless, the types of limitations described in Accounts 1–3 are distinct in principle, and a population's failure could result from limitations in any one.

Account 1 and Account 3 is that RMTS failure on Account 1 result from a lack the capacity to create relational representations that can participate in second-order relational reasoning (i.e., identifying relations between relations), whereas failures on Account 3 result from the lack of relevant learning experiences necessary to have generated such representations despite the presence of the *capacity* to do so.

Let us illustrate the Account 1 versus 3 distinction with a different example: Both a dung beetle and a typical, up-and-coming 9-year-old named Lucky will fail an MTS task in which images of words are to be matched by the words' grammatical class (e.g., Does "run" go with "destroy"' or "and"?). In the case of the dung beetle no amount of experience will lead it to consistently succeed—a permanent failure of the Account 1 type. In contrast, Lucky will initially fail if grammatical classes have not been covered before third grade, but is perfectly capable of succeeding *given the necessary experience*, that is, the requisite grammar lessons—an Account 3 type of failure.

Account 4: Changing Inductive Biases. Possession of a concept does not determine the contexts in which this concept is used. Deciding which of the large repertoire of available concepts one should apply to a given situation is impossible without *inductive biases* that limit the set of possibilities to a manageable few in any given context (see Goodman, 1955). Thus, even if a particular representation (such as *same* or Ω) is available to an individual they may fail to use it in a given context if their inductive biases do not lead them to infer it as relevant. The application to RMTS is transparent: An individual can be perfectly capable of success on the task, already possessing the necessary mental representations and computational capacities, yet will fail if they do not infer sameness and difference to be correct bases of matching. As Livins and Doumas (2015) point out—without recognizing the relations in a stimulus "the rest of the analogy-making process [does] not even get off the ground" (p. 252). The point holds, moreover, if relations are recognized, but not inferred as relevant to the task at hand.

All learning processes that involve hypothesis testing over already represented hypotheses (all Bayesian models) or that involve learning associations between already represented features of the world (all associative models) fall under Account 4. Returning to our example above: An Account 3 failure would, again, be if Lucky did not yet know (have representations of) grammatical classes. In contrast, an Account 4 failure would be correct if Lucky *knew* grammatical classes perfectly well, *and* knew that both "run" and "destroy" are verbs, but her inductive biases lead her to infer that she should match words according to a different property, for example, their approximate length.

Discriminating Among Accounts: Relevance of Training Studies

Accounts 1–4 concern the difference between populations on RMTS performance. If the failure of a given population is due to a capacity limitation (Accounts 1 and 2), no training regime should lead to success on RMTS on the basis of matching same to same and different to different. Clearly, it is always possible that a training regime may induce some strategy that leads to success on a subsequent RMTS task on some basis that other than the relations same and different. If so, this success would not challenge capacity limitation accounts. However, if success on the basis of the relations same and different can be established, training must either have led to new representations of the relations same and different (Account 3) or changed inductive biases so as to increase the likelihood that already existing representations of same and different would be noticed and deemed relevant to the task at hand (Account 4).

We next review a representative sample of training studies that have led to success on RMTS in nonhuman animals. We believe that this literature rules out Accounts 1 and 2 (capacity limitation accounts of failures) for at least some populations that fail RMTS without training. We make no attempt to review every training study. Rather, we review several paradigms with an eye on identifying how existing evidence bears on the various explanations of failures on RMTS by nonhuman animals and young children.

Increasing Salience of Relations

Many training regimes were designed to make the relations same and different more *salient* to a population that otherwise failed at RMTS. Increasing salience is simply changing inductive biases so as to increase attention to a stimulus attribute one already can represent such that it is more likely to be seen as relevant to the task at hand. Thus, if such training is successful, it has changed inductive biases alone, and supports Account 4 of the difference between that population and those who can succeed. One justly famous case study in the literature on RMTS began with exactly this goal: Making the relations same and different more salient by presenting arrays of 16 identical entities (16s) and 16 entities all different from all the others (16d). And indeed, animals who persistently fail standard two-item RMTS succeed robustly at 16-item Array Match to Sample (AMTS), matching 16s to 16s and 16d to 16d. However, in elegant follow-up studies, Wasserman and his colleagues established that the choices were driven by a representation of a *property* of the array, namely, the degree of variability among the entities, or *entropy*—an ensemble statistic like approximate numerosity, or average size of the individuals within the array (see Wasserman & Young, 2010, for review). Thus, success at AMTS does not bear one way or the other on animals' capacity to match on the basis of the *relations* same and different, and does not definitively bear upon adjudicating among the four accounts of failure on RMTS.

Symbol Training

Premack (1983) demonstrated that while chimps without any symbol training failed RMTS, one chimp, Sarah, who had been trained to communicate using a wide variety of plastic symbols, including symbols for same and different, succeeded on the task (replicated by Thompson et al., 1997; training chimps *only* on symbols for same and different). Children's learning or knowing arbitrary symbols for relations (e.g., the words "same" and "different") is also associated with better performance on RMTS (Hochmann et al., 2017), or a partial RMTS task (in which only the relation "same" was used as a sample, see also Footnote 1) in 3- to 4-year-old children (Christie & Gentner, 2014). In addition, after being taught the words "same" and "different," and labels for the dimensions material, color, and shape, a language-trained parrot, Alex, could answer questions such as "how same?," between a pink plastic elephant and a brown plastic giraffe, that is, "material." This was true even though he did not know the words "plastic," pink," "brown," "elephant," or "giraffe" (Pepperberg, 1987, in press). These feats are difficult to square with Accounts 1 and 2. Learning symbols for sameness and difference and using them in tasks with novel stimuli requires generating exactly the kind of abstract relational representation which Penn et al. (2008). argue are impossible for animals to produce (Account 1), and requires the necessary computational capacities to manipulate these representations (Account 2).

Success after symbol training is consistent with both learning history accounts: Parrots, chimpanzees, and young children are clearly capable of generating symbolic representations of sameness and difference. Thus, symbol training may have led them to generate these representations for the first time (Account 3). It is also possible that these groups already had mental representations of the relations in a format that could easily be mapped onto an external symbol (a summary symbol *same* or Ω). On this view training simply mapped these existing mental representations to external symbols, in the process changing inductive biases so as to make these relations more salient and likely to be inferred as meaningful bases of response in a task (Account 4).

Dogged Training

Some of the strongest evidence that nonhuman primates have representations of the relations same and different sufficient to support RMTS derives from training regimes Premack(1983) deemed "dogged training" (see Wasserman et al., 2017, for review). These studies provide extensive correct/incorrect feedback on RMTS, up to 60,000 trials. For instance, after 17 to 30 thousand trials, a minority of baboons (six out of 29 individuals) performed above chance on RMTS (Fagot & Thompson, 2011; see also Truppa et al., 2011, for similar findings in capuchin monkeys).

Some concerns remain as to whether such successes are truly a result of animals responding on the basis of abstract relational representations: For instance, many animals who reach above-chance levels on one set of RMTS stimuli after dogged training fall to chance with novel stimuli, suggesting that in tens of thousands of trials, which included repetitions of stimulus triads, they had learned the correct responses for a subset of particular stimuli (see Wasserman et al., 2017, for review). Nonetheless, a small minority of baboon and capuchin participants do succeed on transfer trials with entirely novel stimuli. These successes as a result of dogged training are hard to square with capacity accounts (1 and 2) since there is no obvious way in which correct/incorrect feedback alone would produce success via nonrelational strategies. In contrast, such training may have led animals to generate new representations of sameness and difference as a result of repeated comparison of sample and choice pairs (Account 3, see e.g., Gentner & Hoyos, 2017 for the argument that comparison facilitates abstraction of new relations). Likewise, dogged training may have led them to infer preexisting representations of sameness and difference as correct bases of matching as a result of testing and rejecting an enormous number of alternative possible bases of matching (Account 4).

Progressive Alignment

A training paradigm known as "progressive alignment" (Kotovsky & Gentner, 1996) has been shown to induce relational matching in preschool children, in the face of failure on the same tasks in the absence of progressive alignment. This paradigm first presents individuals with matches which are *both* matches on object properties *and* matches on relations (e.g., in the case of RMTS, matching AA to AA and not BC). After some number of such trials, the object matches are removed, leaving a purely relational matching task (e.g., matching CC to DD and not EF, as in standard RMTS; Figure 1). Success after progressive alignment training transfers to novel stimuli and would be incompatible with Accounts 1 or 2. There is no obvious way in which progressive alignment could have led children to succeed via nonrelational strategies.

Gentner and her colleagues propose that progressive alignment produces new relational representations (Account 3)—for example, "analogical comparison [including progressive alignment] is . . . the main driver of *new relational abstractions*" (Gentner & Hoyos, 2017, p. 687, emphasis added, see also Ferry et al., 2015; Kotovsky & Gentner, 1996). This is likely the case when the product of progressive alignment is a complex relation unlikely to have been formulated before (e.g., black figure above white figure, Christie & Gentner, 2010), or is a relation encoded by a novel verb (e.g., a verb meaning "to hold behind your back and then put down," Haryu et al., 2011). Progressive alignment might also facilitate the construction de novo of a new summary symbol Ω that is coined to represent what is in common among AA, BB, CC, DD—namely sameness. It is also possible, however, that progressive alignment draws attention to a preexisting representation of sameness and difference (i.e., same or Ω already was in the animal or child's repertoire) by their constant co-occurrence with reinforced object matches (Account 4).

Smirnova et al. (2015), Obozova et al. (2015)

The present studies are motivated by the stunning success of two crows (Smirnova et al., 2015) and two parrots (Obozova et al., 2015) on *three separate* RMTS tasks after a complex, two-part training paradigm. In the first part of the paradigm, experimentally naive birds where trained to criterion on a series of MTS tasks: Identity/Color MTS (matches on all dimensions, mismatches on color), Identity/shape MTS (matches on all dimensions, mismatches on shape), and Number MTS (matches on the number of objects per card). They were also tested on blocks of trials in which they needed to flexibly shift between matching on color identity, shape identity, or number, including on new values along those dimensions, different from those in the initial training. The birds succeeded within the first testing block on these mixed trials. In Smirnova et al. (in press), the authors suggest that birds had learned the logic of matching tasks in this training—to identify dimensions on which the choice cards differ and on which the sample matches only one of the choice cards, and furthermore could flexibly shift among different properties that satisfy the logic of matching tasks.

They then put this hypothesis to a strong test, seeing whether birds would generalize to matching a *previously untrained* object dimension, size. The Size MTS task was composed differently than previous MTS tasks: Birds completed sets of four trials in which three reinforced progressive alignment training trials where objects matched not only on size but on all features (e.g., C goes with b or C) were followed by one nondifferentially reinforced test trial in which objects matched only on approximate size (e.g., x goes with y or Z). Birds succeeded on the nondifferentially reinforced Size MTS test trials from the very first training session.

After Size MTS, animals were tested on three RMTS tasks: Size RMTS (same size goes with size; different size goes with different size), Color RMTS (same color goes with same color, different color goes with different color), and Shape RMTS (same shape goes with same shape, different shape goes with different shape). The structure of RMTS tasks was the same as Size MTS—sets of four trials with three reinforced progressive alignment trials (e.g., AA goes with AA or BC) followed by one nondifferentially reinforced test trial (e.g., DD goes with EE or FG). Both crows and parrots succeed robustly on RMTS test trials from the first session with an average of 83.33% and 72.22% correct for crows and parrots, respectively. Notably, this was on par with their performance on reinforced progressive alignment trials within the same session—76.39% and 75%, crows and parrots, respectively—on which matches were *both* on relations *and* on object features.

Interpreting Birds' Success on RMTS. Crows and parrots clearly have the representational and computational capacities to succeed on RMTS. Their success on the probe test trials is "spontaneous," in the sense of being from the first session of test trials and in the absence of error feedback on test trials, and unprecedented in the animal literature. However, it is not spontaneous in the sense of "untrained." The differentially reinforced progressive alignment training on three-quarters of trials were, in effect, training trials. It is not known whether birds' progressive alignment training was necessary, or even sufficient, for success on the crucial test trials—and if so what the effect of progressive alignment was. If the progressive alignment trials *were* necessary, and *if* progressive alignment leads to new abstract mental representations of the relations same and different, these results are consistent with Account 3—that is, that differences between untrained crows and parrots, on the one hand, and human adults, on the other is in the absence versus presence of representations of sameness/difference with the right properties to support RMTS (e.g., a summary symbol like *same* or Ω).

That being said, the equivalent proportion of successful matches on full RMTS test trials compared with progressive alignment trials by birds even in the very first test session lend weight to the possibility that progressive alignment was not the only process involved in birds' success. The other possible contribution to birds' remarkable success on RMTS test trials is their previous training on the series of MTS tasks described above. This possibility is made all the more remarkable by the fact that, while at least some (and according to a reviewer on Kroupin & Carey, 2020, *most if not all*) other comparative studies included MTS training before test, these did *not* facilitate the use of relations in a subsequent matching task (e.g., MTS-trained baboons in Fagot et al., 2001, went on to solve AMTS using entropy representations and *not* the relation same).

Clearly, MTS training cannot alleviate absolute representational or computational capacity limitations (Accounts 1 and 2). The possible effects of MTS in terms of Accounts 3 and 4 depend on whether Hochmann et al. (2016) and Zentall et al. (2018) are correct in saying MTS involves only a match computation and *not* a mental symbol *same*. If MTS does involve a representation *same* (contra Hochmann et al., 2016, and Zentall et al., 2018), Account 3 cannot be correct since, ex hypothesi, success on MTS would imply the availability of the relational representations required to succeed on RMTS. If MTS does *not* involve *same*, Account 3 would have to assume that birds initially succeeded on MTS tasks without the representation *same*, then after success had already been achieved this representation emerged de novo via a mechanism we have tried and failed to imagine. This already obscure (to us) possibility is further complicated by the apparent evidence that not all MTS training is sufficient to produce such a representation (e.g., in the case of MTS training not facilitating relational responses in Fagot et al., 2001). In summary, if MTS training tasks increase second-order relational responding, we see no obvious way in which this is consistent with Account 3—that the effect of training is to produce new relational representations.

The possible effects of MTS training on Account 4 are more straightforward: If MTS does involve the representation *same*, then training on *any* MTS task should change inductive biases so as to make this representation to be more likely to be used in a subsequent RMTS task. Once again, the evidence that not all MTS training tasks facilitate RMTS success weighs against this possibility. If MTS does *not* involve the representation *same*, then some other properties of the MTS tasks must affect individuals' inductive biases so as to make *same* relatively more likely to be inferred as the correct basis of matching. On this latter possibility, it need not be the case that *all* MTS training tasks facilitate relational responding.

Interim Conclusion and the Present Studies. Despite the possibility of success via nonrelational strategies in some cases, existing evidence overwhelmingly weighs against capacity limitation accounts (Accounts 1 and 2) of crows', parrots', monkeys', and apes' failures on RMTS. Previous work cannot, however, distinguish between Accounts 3 and 4 since all previous training regimes have involved either reinforced second-order relational matching trials (e.g., dogged training, progressive alignment) or explicit symbol training.

The possibility that the MTS training in the parrot and crow studies played a necessary, perhaps even *sufficient*, role in the birds' success provides a wedge into beginning to distinguish Accounts 3 and 4 since they involve neither second-order matches nor symbol training. If MTS training tasks can be shown to increase second-order relational responding *without* the possibility of supporting new specific representations of the relations same and different between pairs of individuals, this would support the plausibility of an inductive bias account of at least some population differences (Account 4). The **first** aim of the present experiments is to test this hypothesis: We *ensure* that new representations of the relations same and differ are not produced by MTS training tasks by testing human adults, who manifestly *already have* the requisite, fully abstract, explicit representations of these relations. Our **second** aim is to determine whether or not *any* MTS task is sufficient to facilitate second-order relational responding (that would be consistent with MTS involving *same*), which we do by testing the effects of MTS training tasks one at a time. **Third**, we begin to explore the *mechanisms by which* MTS training may increase second-order relational responding.

Experiment 1

Method

The logic of our investigation requires a dependent variable which could reflect increases in the tendency to match on relations after MTS training tasks. Clearly, RMTS is the perfect candidate in populations that have been shown to fail on the task. To our knowledge, however, there are no published data regarding adults' performance on a RMTS task where both the sample and choice cards display two items. Experiment 1 tests adults' performance on standard RMTS to establish whether adults choose incorrectly on a sufficient proportion of trials such that we could potentially see an *increase* in relational matching as a result of completing MTS training tasks.

The possibility that adults may make a significant number of incorrect choices is supported by results from a paradigm where adults' use of same/different representations were assessed using two-item arrays—the same/different discrimination paradigm. Participants saw one pair of items on the screen at a time and had to learn to press one of two buttons as a function of the relation instantiated by the pair of stimuli (e.g., if same, press left; if different, press right). Strikingly, adults found this task quite difficult. For example in one study, after 48 training trials with feedback as to whether the choice was correct or incorrect on every trial, 52% of college students failed to achieve criterion of 70% or more correct across a block of 12 trials (Castro & Wasserman, 2013). Given this striking failure of a *majority* of adults on a task requiring responding on the basis of the relations same and different, adults' spontaneous (no error feedback) success on RMTS is hardly a foregone conclusion.

Participants

We recruited participants via Amazon Mechanical Turk (MTurk). Participants were 601 adults who had not participated in any of our MTurk RMTS studies in the past (M_{age} = 34.65, SD = 10.36). Participants were recruited from the United States only and were given a small monetary compensation for participating. Recruitment and compensation policies were the same across all experiments reported here.

Procedure

All procedures and materials in all experiments reported here were approved by the Harvard University Institutional Review Board (IRB), and all participants gave informed consent before the beginning of each study. Participants first saw several instruction screens indicating that they would be completing a matching task in which one of two cards on the bottom of the screen would match the card on the top of the screen. After this introduction, participants completed eight RMTS test trials (see Figure 1, e.g., of stimulus triads, see Appendix for full description of the stimuli). The appearance of the trial screens was designed to mimic non-computerized RMTS paradigms such as Smirnova et al. (2015): Pairs of geometric figures were enclosed in "cards" (i.e., thin black rectangles). On four of the trials the card at the top of the screen instantiated the relation same, and on four it instantiated the relation different. On each screen one card at the bottom of the screen instantiated the relation same, the other different. The left/right position of the correct choice was fully counterbalanced. Each screen displayed the prompt "Which one of the two sample shapes goes with the target above?"[3]

All objects in the task were unique, with the exception of those repeated within each same card for a total of 36 distinct objects across the task. Figures differed in shape and color and were equal in height and width. Figures were placed on the middle of the cards' vertical axes, and equally spaced from the horizontal axis. Each set of six figures (two on each of three cards) always appeared together in a trial and each set of six appeared only once during the experiment. The order of trials was randomized. Participants selected which of the two bottom cards they believed went with the top card by clicking on a button below the respective image. After selecting one of the options participants advanced to the next trial. *Participants received no feedback on their performance at any point during the task.* After participants completed all eight trials of RMTS, they saw a screen thanking them for participating in the study and were asked to indicate their age.

Results

Adults were overwhelmingly successful in two-item RMTS, choosing the relational match on 96% of total trials. Overall, 82% of participants succeeded from the very first trial, choosing correctly on all eight trials. An additional 11% of the participants made seven of eight relational responses (statistically above-chance performance, binomial test, p = .04). Because there was no error feedback on this task, this means that almost all (93%) individually succeeded above chance, that is, succeeded spontaneously, with no feedback.

Clearly, adults' performance on RMTS is too close to ceiling for us to use RMTS as a dependent variable in testing the effects of MTS training tasks on subsequent relational matching. Furthermore, such high levels of spontaneous success stands in stark contrast to the failure of 52% of adults on same-different discrimination despite extensive training (Castro & Wasserman, 2013). A number of factors may account for this difference: First, an understanding of the rules of the matching task involve comparing top and bottom cards in search of a salient dimension on the basis of which the top card is similar to *just one* of the bottom cards (and deciding which basis of similarity is most likely to be correct if more than one is identified). We have referred to this as "understanding the logic of matching tasks." The logic of matching tasks provides constraints on the potentially correct bases of matching that are not present in same-different discrimination tasks. Second, the logic of matching tasks involves comparing cards, and comparison has been shown to promote the salience of relations (e.g., Markman & Gentner, 1993). Third, the figures in the Castro and Wasserman (2013) stimuli were relatively semantically rich—pictures of dice, books, cameras, and so forth—and were located in various positions on the displays. In contrast, the figures in our study were colored geometric shapes that were always side by side in the center of the card (see Figure 1). Participants in Castro and Wasserman (2013); therefore, may have been more likely to attribute complex interpretations of the stimuli based on their semantic content and relative positions—distracting from basic same/different relations as a hypothesized basis of responding. Examples of the rules verbalized by participants who failed same-different discrimination provides anecdotal evidence that at least some failures were due to generating complex semantic interpretations (Castro & Wasserman, 2016).

---

[3] The term "sample" in the context of (R)MTS tasks has historically been used to refer to the singleton card, not the two choice cards. Here we use the term to refer to what are more commonly called the "choice" cards, that is, those from among which the match is *selected*. To avoid confusion we will in the body of the paper refer to the matched-to and selected-from cards (sample and choice cards in traditional terms) as "top" cards and "bottom" cards, respectively. We discuss the use of the term "sample shapes" in Footnote 5.

In summary, Experiment 1 demonstrates that adults are not only capable of representing and making matches on the relations same and different but that they do so spontaneously in a task that involves comparisons between sets of colored geometric shapes.

Experiment 2

Method

Adults' ceiling-level performance on RMTS leaves little room to change their preexisting inductive biases such that they become *more* likely to make relational matches. Thus, we require a task which contains *both* a relational match *and* another basis of matching, (a paradigm called "cross-mapping" by Rattermann & Gentner, 1998), such that a significant portion of adults do *not* match on relations despite clearly being *able* to do so (as evidenced by Experiment 1).

Previous work with adults has provided two such models relevant to RMTS: First, Vendetti et al. (2014) have shown that adults' choices were evenly split (50%) in a task that instantiates the choice between two types of matches: A match between two objects different in appearance but involved in the same relation (a relational match) and a match between two objects identical in appearance on all dimensions but not involved in the same relation. Second, Christie & Gentner (2007) gave adults a simplified RMTS task in which the top card always displayed the relation same and one of the two bottom cards showed a relational match (i.e., two identical shapes both of which differed in shape and color from the sample shapes) as in standard RMTS. In the other bottom card, one of the two objects was identical to the objects on the top card in shape but not in color; the other object differed in from the sample shapes on both.[4] We will refer to this latter kind of match as an incomplete partial object match. An incomplete object match is one in which one but not both of the objects on the choice card match the objects on the sample card. A partial object match is one in which an object on the choice card matches the objects on the sample card on some but not all dimensions. Though adults in the Christie and Gentner task were somewhat more likely to make relational matches than incomplete partial object matches there was still a sizable proportion of trials (31%) on which adults did *not* choose relational matches. This is dramatically more than the tiny minority of trials (4%) in which adults did not choose relational matches in RMTS in Experiment 1, but noticeably less than the proportion (50%) of trials on which adults chose object matches (that were not partial) in Vendetti et al. (2014).

Synthesizing these findings, in Experiment 2 we developed a modified RMTS task aimed at retaining the structure of RMTS while maximizing the likelihood that adults would choose a nonrelational basis of matching *over* a relational one. Specifically, our new task ("Object Match-to-Sample vs. Relational Match-to-Sample," OMTSvRMTS) is identical to that of Christie and Gentner, except that the nonrelational matching card was an incomplete, *but not partial*, object match. That is, the top card always displays the relation same, one bottom card displays a relational match alone (i.e., two identical figures, each different from those on the top card) and the second bottom card displays two different objects, one of which was identical to the objects on the top card *on all dimensions*—that is, an incomplete object match—and the other differed from the objects on the top card both in shape and color (see Figure 2 and Appendix for full description of the stimuli).

Participants

We recruited 193 adult participants from MTurk who had not participated in any RMTS study from our lab ($M_{\text{age}} = 37.31$, $SD = 11.46$).

Procedure

The procedure for Experiment 2 was identical to that of Experiment 1 except participants completed eight trials of OMTSvRMTS instead of RMTS, receiving no feedback as to correctness of their choices.[5]

OMTSvRMTS: The left/right position of relational and incomplete object match cards was counterbalanced. Likewise the position of objects on the incomplete object match cards was counterbalanced such that the object identical to the objects on the top card appeared on the right side of the card in half of the trials and on the left in the other half (compare the two incomplete object match cards in Figure 2). Each trial contained three unique objects and no objects were repeated across trials for a total of 24 unique objects in the task.

After participants completed eight trials of OMTSvRMTS, they saw a screen thanking them for participating in the study and were asked to indicate their age.

Results

Experiment 2 generalizes to new stimuli the finding that, faced with a task that pits a relational match against an incomplete object match, adults were roughly evenly split as to which basis of matching they chose: 44% of all responses were incomplete object matches and 56% were relational matches. This distribution was not significantly different from an even split (independent sample *t* test, $t(792) = 1.94$, $p = .053$, two-tailed).

Furthermore, the even split between object and relational matches did not reflect an even split *within* responses by each individual participant. Rather, the participants in Experiment 2 overwhelmingly settled on a consistent basis of response. We use seven or more out of eight choices as a criterion for consistent responding as this number of responses of one type is significant under a binomial test ($p = .04$). Under this criterion, 54% of participants consistently preferred relational matches and 41%

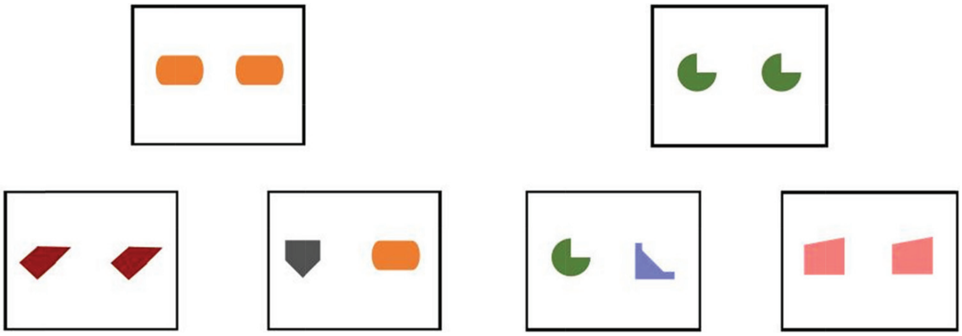
---

[4] This is also the same structure as the generalization task used by Fagot and Thompson (2011) with baboons after tens of thousands of training trials on standard RMTS. While the structure of the task is the same, the extensive RMTS training in Fagot and Thompson (2011) makes it a test of *generalization* (of relational matching). In contrast, since adults were not trained to match on relations the tasks discussed here are measures of *spontaneous performance*.

[5] In retrospect, it would have been better to ask "which of the two sample *cards* goes with the target above" rather than "which of the two sample *shapes* . . . ," because the latter locution may invite attention to individual figures rather than to the relation. To allay this concern we replicated Experiment 2, replacing "sample shapes" with "sample cards." The results were identical. In any case, given that the focus of this work are potential differences *between* experimental conditions, the current results cannot be attributed to details of the wording as it was identical throughout. Therefore, we did not replicate the rest of the experiments using the term sample cards.

Figure 2

Two Object Match-to-Sample Versus Relational Match-to-Sample (OMTSvRMTS) Trials



Note. OMTSvRMTS = Object Match-to-Sample vs. Relational Match-to-Sample. In the left triad, the relational match is the card on the left and the incomplete object match is the card on the right. In the triad on the right, the incomplete object match is the card on the left and the relational match is the card on the right. See the online article for the color version of this figure.

consistently preferred incomplete object matches. This high degree of consistent responding (95% of all participants) suggests that adults have a strong bias to use the same rule on all trials, at least in the case of this matching task.

In summary, Experiment 2 confirms that a sufficiently large proportion of adults, while clearly *able* to make relational matches (Experiment 1) do *not* do so in OMTSvRMTS. Thus, OMTSvRMTS offers the opportunity to explore whether MTS training tasks might measurably *increase* the likelihood of relational matching in adults.

Experiment 3

Method

Experiment 3 addresses our first two aims: **First**, it tests whether training which *could not have* led to the first abstract representations of the relations same and different can increase relational responding by changing inductive biases *alone*. Specifically, we test the hypothesis that successfully completing at least some MTS tasks adapted from Smirnova et al. (2015) and Obozova et al. (2015) will make human adults more likely to make relational matches in OMTSvRMTS than at baseline (Experiment 2). Eight trials of MTS training cannot change the nature of human adults' representations of the relations same and different so as to allow them to support RMTS. This is because adults demonstrably already have representations that can do so, as confirmed in Experiment 1. **Second**, it tests whether such effects, if observed, are due to relational content within the MTS task itself. If so, training on all four MTS tasks should increase in relational responding on OMTSvRMTS.

In the studies below we do not aim to emulate the *process* of training in Smirnova et al. (2015) with adults but rather its *end result*—that is, the successful completion of the MTS tasks. Crows in Smirnova et al. (2015) succeeded (eventually) on all MTS tasks before they were tested on RMTS. Consequently, in the studies below we will consider the effects of MTS tasks on OMTSvRMTS *only* for those adults that succeed above chance (seven or more out of eight trials correct) on the former. With this in mind, we strove to maximize the proportion of adults succeeding on each MTS task while minimizing the extent of training. This was ensured by instructing adults as to the correct basis of matching if they made an incorrect choice on the MTS training task.[6] There was no feedback on the subsequent OMTSvRMTS test task.

Even with such corrective instructions, adults' performance on MTS tasks can provide valuable information for beginning to explore the *mechanisms by which* training can change inductive biases so as to increase relational responding. Such an exploration must start by establishing what adults' pattern of inductive biases *is* in the first place, before any training. The proportion of adults succeeding spontaneously (i.e., choosing correctly on all trials and receiving no instruction) on a given MTS task indicates how strongly their preexisting inductive biases align with the relevant basis of matching (e.g., matching on color in Color MTS). In fact we already have such data for the adults' inductive biases regarding sameness and difference—the rate of 8/8 trials correct in Experiment 1 was 82% (see Figure 3). Combining these data regarding adults' pattern of preexisting inductive biases with the effects of MTS training tasks on rates of relational responding in OMTSvRMTS will allow us to generate hypotheses as to the *mechanisms* by which MTS training may change adults' original inductive biases so as to increase relational responding.

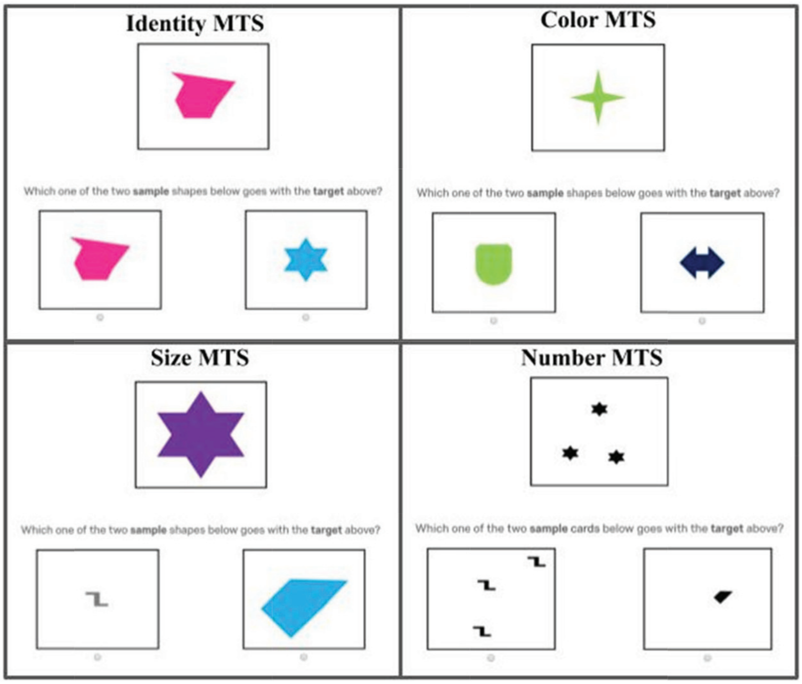
Participants

Participants were recruited via MTurk as in Experiments 1 and 2. None had participated in any previous (OMTSv)RMTS study in our lab. Each of Experiments 3A–D had two sample sizes. First, the total number of participants who completed the task was used to analyze spontaneous success rates on MTS tasks. Second, because we are interested in the effects of *successful* MTS training on relational responding, only those participants who succeeded above chance on MTS (i.e., made at most one mistake on eight

---

[6] Providing instructions on correction also allowed us to replicate the training tasks from Experiment 3 exactly with young children to see if the same training would lead to success on RMTS at an age where children otherwise fail (Kroupin & Carey, 2020). To establish that the critical results reported here would be the same even if we did not tell participants the criterion of matching on the training MTS tasks if they erred, we repeated Experiment 3 with Size MTS and Number MTS training having error feedback alone (i.e., participants were only told that they chose incorrectly and not given instructions). The pattern of results was identical.

Figure 3
Example Trials of Match-to-Sample (MTS) Tasks in Experiments 3A–D



Note. See the online article for the color version of this figure.

trials) were used in analyzing the effects of MTS training on OMTSvRMTS. In all cases the number of participants excluded under this criterion was minimal, and including them in no way changes the pattern of results. For Experiments 3A–D the total N (and N of MTS-succeeders) was as follows. 3A: 183(181); 3B: 194(186); 3C: 180(173); 3D: 192(170). The mean ages(and standard deviations) of the total sample sizes were as follows. 3A: 34.47(11.23); 3B: 35.57(11.61); 3C: 35.54(10.98); 3D: 35.72 (11.21).

General Design

All of Experiments 3A–D were between-subjects, that is, any given participant completed only one MTS task and was tested only once on OMTSvRMTS. Each participant completed eight trials of one MTS training task: Identity MTS (Experiment 3A), Color MTS (Experiment 3B), Number MTS (Experiment 3C), or Size MTS (Experiment 3D; see Figure 4) followed by the OMTSvRMTS test task of Experiment 2. While the *bases* of matching were similar between our MTS tasks and those used by Smirnova et al. (2015; i.e., identity, color, or number, size), the stimuli in Experiment 3 differed in many respects from the MTS tasks used by Smirnova et al. (2015; see Appendix and Smirnova et al., in press; for detailed descriptions of the stimuli in the respective paradigms). These differences, however, do not affect the hypothesis being tested in Experiment 3—namely that training on simple MTS tasks can increase relational responding on a subsequent RMTS in a population that demonstrably already has abstract representations same and different sufficient for supporting RMTS.

Participants were given explicit instructions as to the correct basis of matching if they chose incorrectly in an MTS training task. For instance, if a participant chose incorrectly on Size MTS, they saw a screen with the following text: “Good guess! But that was the wrong choice. In this game cards with big shapes go with cards with big shapes and cards with small shapes go with cards with small shapes.” After completing the MTS training task, participants completed eight trials of OMTSvRMTS, identical to Experiment 2, with no feedback whatsoever. Details of the stimuli and procedures for each MTS training task can be found in the Appendix.

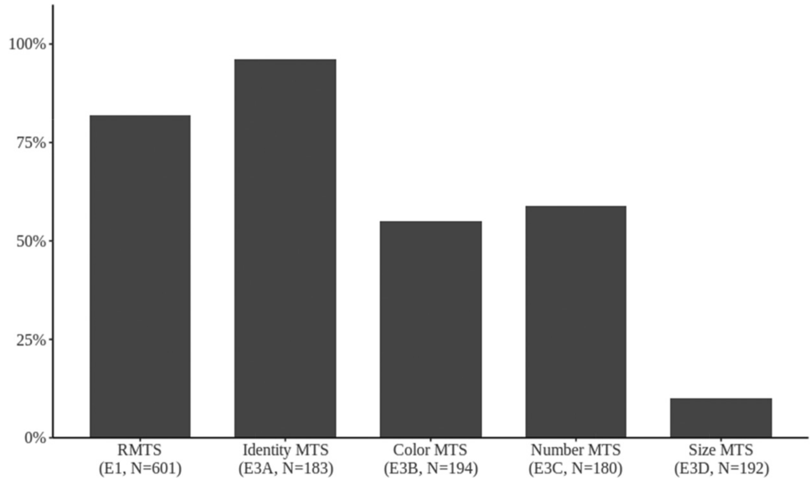
Results

MTS Training Tasks

The proportion of adults spontaneously succeeding on each MTS training task (i.e., choosing eight out of eight times correctly and receiving no correction) are displayed in Figure 3. The proportion succeeding spontaneously on RMTS in Experiment 1 is also included for comparison. Note that for RMTS in Experiment 1, because there was no error feedback, 7/8 correct was statistically spontaneous success. However, for comparability to Experiment 3 where participants were told the correct basis of matching if they made one error on a MTS task, we took 8/8 correct on RMTS, as well as on each MTS task, as our criterion for spontaneous success

Figure 4

Percentage of Adults Spontaneously Succeeding (8/8 Trials Correct, No Feedback) On (Relational) Match-to-Sample ((R)MTS) Tasks in Experiments 1 and 3A–D (Tasks and Sample Sizes Are Indicated on the x Axis Labels)



in these analyses. The proportion of spontaneous succeeders differed across tasks $\chi^2(4, N = 1{,}352) = 426.57$, $p < .0001$. Post hoc tests with Bonferroni correction indicated that all comparisons between proportions of spontaneous succeeders across tasks were significant, with the exception of the proportions in Color and Number MTS, which did not differ from each other.

Clearly, matching on object identity is highly in line with participants' preexisting inductive biases (96% spontaneous success), more so than matching on the relations same and different (in RMTS, 82%), while matching on color and number is markedly less so than either of the first two bases (55% and 59%, respectively). The rate of spontaneous success at matching on the basis of object size was *strikingly* low (10%), suggesting the preexisting inductive biases of adults in our sample are highly unlikely to lead them to match geometric figures based on their size. In part, this may be a result of the fact that geometric shapes do not have canonical natural size. This is in contrast to adults' and children's perceptions of naturalistic stimuli, such as a picture of a house or a lamp, whose real-world sizes are computed automatically by the visual system (e.g., Long & Konkle, 2017; Long et al., 2019).

Because participants were *told* the intended basis of matching if they made an error on one of the MTS training tasks, it is not surprising that most participants made at most one error; thus, succeeding above-chance statistically on each training task. The proportion of adults who made at most one error was 98% on Identity MTS, 96% on Number MTS, 96% on Color MTS, and 88% on Size MTS. The proportion of participants performing above chance differed by MTS task $\chi^2(3, N = 750) = 20.57$, $p < .001$). Post hoc analyses with Bonferroni correction revealed that Size MTS was the outlier—the proportion of above-chance succeeders on Size MTS was significantly different from the proportion in all other MTS tasks, while the proportions among Identity, Color and Number MTS did not differ from each other. This result reinforces the conclusion that matching geometric figures by their size is contrary to adults' preexisting inductive biases—so much so that 12% of participants continued to make errors *even after* being told that they should match big figures to big figures and small figures to small figures.

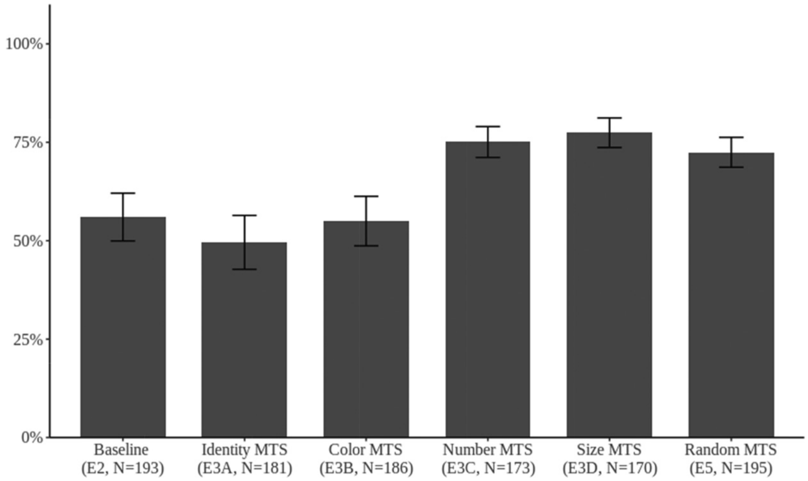
OMTSvRMTS

All participants were included in the analyses reported above of performance on the MTS training tasks. In testing the hypothesis that successful MTS training would affect performance on the subsequent OMTSvRMTS task, we removed the small minority of participants who did not succeed even after correction on the training MTS tasks (i.e., made more than one error on the task) when analyzing OMTSvRMTS performance. The pattern of results (both here and in Experiment 5), however, remains unchanged if OMTSvRMTS data from these participants are included in the analysis.

Participants received no feedback on OMTSvRMTS, which always followed the MTS training task in Experiment 3. They were free to match on the basis of the relation same or on the basis of an incomplete, but perfect, object match (see Figure 2). As in Experiment 2, individual participants were overwhelmingly consistent in choosing either incomplete object matches *or* relational matches in seven or more of the eight trials OMTSvRMTS. In no experiment did the proportion of consistent choosers fall below 86%. The dependent variable we explore to establish the effects of training tasks is the percentage of relational matches across all eight trials of OMTSvRMTS. Figure 5 displays the percentage of relational matches in Experiments 3A–D as well as in the no-training baseline (Experiment 2) and in a study reported below that tested a hypothesis concerning a mechanism through which MTS training might affect subsequent performance (Experiment 5). An analysis of variance (ANOVA) examined the percentage of relational matches in OMTSvRMTS across Experiment 2 (baseline) and the four MTS training conditions of Experiment 3 (Color, Identity, Number, and Size). There was a main effect of training condition (No training, Identity MTS, Color MTS, Number MTS,

Figure 5

Percentage Relational Matches on Object Match-to-Sample Versus Relational Match-to-Sample (OMTSvRMTS) With No Training (Baseline, Experiment 2), and After Training On Identity MTS, Color MTS, Number MTS, Size MTS, and Random MTS (Experiments 3 and 5)



Note. Sample size excludes participants who did not succeed above chance on MTS training. Error bars display standard errors.

and Size MTS; ($F(4, 898) = 14.80$, $p < .0001$). Post hoc tests using Tukey's honest significant difference (HSD) criterion revealed that the proportion of relational matches in OMTSvRMTS did not differ across baseline (Experiment 2), and after Identity and Color MTS training (Experiment 3). Likewise, the proportion of relational matches did not differ between conditions training on Number and Size MTS (Experiment 3)—both of which were significantly higher than all of baseline, Identity and Color MTS conditions. In other words, while training on Identity and Color MTS did not change the percentage of relational matches made by adults on OMTSvRMTS, a mere *eight trials* of training on Number or Size MTS significantly increased the likelihood of adults engaging in relational reasoning on OMTSvRMTS.

Interim Discussion

Effects of MTS Training

Experiment 3 yielded two important results: First *eight trials* of Number and Size MTS training increased adults' spontaneous second-order relational responding. Second, the same number of trials of Identity and Color MTS training *did not*. We discuss the results in order.

Our **first** goal was to test whether MTS training tasks of the type used by Smirnova et al. (2015), can change inductive biases so as to increase the likelihood of second-order relational reasoning on a subsequent task. Experiment 3 provides striking evidence that they can: A mere eight trials of Number *or* Size MTS significantly increased adults' second-order relational responding on a subsequent OMTSvRMTS task. Given adults clearly have the representations and computational capacities required to match on same/different relations (Experiment 1) the effects of MTS training *can only have been* to change their inductive biases.[7] There is an extensive literature showing increased relational responding in adults as a result of experience that drew attention to relational content (e.g., Gick & Holyoak, 1980; Jamrozik & Gentner, 2020; Vendetti et al., 2014). However, to our knowledge, this is the first demonstration of a facilitation of relational matching via a training experience (MTS tasks) that involved no second-order relational reasoning.

Our **second** goal was to test whether training on *any* MTS task would facilitate second-order relational matching in OMTSvRMTS. Such a finding would be consistent with the possibility that MTS *is* completed in terms of a rule that involves an abstract symbol for same encoding the relation between the sample and the correct choice, despite such a symbol not being *necessary* for MTS success, as discussed above. Experiment 3 provides clear evidence that it is not the case that any MTS training affects the likelihood of relational matches. While two MTS tasks (Number, Size MTS) did increase adults' second-order relational responding in OMTSvRMTS, the other two (Identity, Color MTS) did not. Thus, even for a population which clearly has the representation *same* readily available (as evidenced in adults' robust success on RMTS in Experiment 1) this representation does not seem to be involved in solving MTS tasks.

These results also help make sense of an otherwise puzzling pattern of findings in the comparative literature alluded to above: One the one hand, some studies find that Identity MTS training does not ensure success at RMTS (e.g., Fagot et al., 2001, and according to a reviewer for Kroupin & Carey, 2020, *most* nonhuman animals in RMTS studies have had previous experience with Identity MTS). On the other hand, there is reason to suppose the MTS

---

[7] Note that while we are sure adults did not generate new same/different representations, this does *not* mean that participants constructed no new representations *at all*. Doubtless they generated some new representations—of the novel stimuli involved in the task, at least. The critical issue in understanding the role of training, however, is not whether *any* new representations were generated, but the fact that no new representations of the abstract same/different relations involved in RMTS were generated.

training in Smirnova et al. (2015) and Obozova et al. (2015). played an important role in birds' RMTS success: Even though there were progressive alignment trials, birds succeeded on the nondifferentially reinforced RMTS test trials from the very first session, performing at the same level as on the two types of trials. The results of Experiment 3 suggest that the *kind* of MTS task is crucially important to facilitating second-order responding— consistent with the fact that, as far as we can ascertain, Smirnova et al. (2015) and Obozova et al. (2015) were the first to train animals on MTS tasks *other than* Identity, Shape, or Color MTS, before testing on RMTS.

Characterizing Inductive Biases and Explaining Changes in Inductive Biases

Having established that at least some MTS training tasks can increase second-order relational responding by changing inductive biases alone, we turn to the **third** aim of these studies: Beginning to explore these inductive biases and the mechanisms through which training might change them. Specifically, we seek an account of our two major findings: Why Number and Size MTS training tasks increased subsequent relational reasoning, whereas Identity and Color MTS did not change the percentages of relational matches, relative to baseline.

The goal of exploring such inductive bias mechanisms is shared with other recent work that explores how training changes inductive biases so as to promote relational reasoning (e.g., Simms & Richland, 2019; Vendetti et al., 2014). These studies gave some participants (adults, Vendetti et al., 2014; children, Simms & Richland, 2019) experience with completing analogies—a second-order relational reasoning task—before testing all participants on a matching task with unrelated stimuli and relations. The matching test task, like OMTSvRMTS, contained both relational and object-feature matches. Participants who received second-order relational experience were found to be more likely to make relational responses in the subsequent matching task. Both Vendetti et al. and Simms & Richland concluded that relational responding was increased as a result of second-order relational training changing inductive biases so as to increase attention to relations *in general*—facilitating a "relational mind-set."

This work leaves the mechanisms by which the facilitation of relational responding occurred largely unspecified, with Simms and Richland suggesting only that "Our findings, along with those of Vendetti et al. (2014) are consistent with the idea that once effortful [second-order] relational processing is engaged, its momentum can carry forward to new situations" (p. 10). The results of Experiment 3 complicate this picture for at least two reasons. First, both Number and Size MTS increased second-order relational processing in OMTSvRMTS despite not being second-order tasks. Second, only *some* MTS tasks had this effect (i.e., Number and Size but not Identity or Color MTS). Clearly, we need to develop more detailed understanding of the mechanisms involved in inductive-bias change relevant to second-order relational reasoning to account for the results of Experiment 3. We propose that this can be achieved via analysis at the level of *inductive biases over specific representations*, that is biases to match on the basis of specific properties of entities or specific relations among them, in contrast to degrees of domain-general preference for relational matches. This level of analysis involves specifying two things: First, the relevant preexisting inductive biases of the population relevant to the task at hand - here U.S. adults and OMTSvRMTS, respectively. Second, the mechanisms by which these biases are changed by training—here MTS tasks in Experiment 3.

Preexisting Inductive Biases. The pattern of spontaneous success on RMTS (Experiment 1) and MTS tasks (Experiment 3; see Figure 3) show adults' preexisting inductive biases make them likely to infer matches on the relation same (the basis of correct responding on RMTS) as correct, as well as combined matches on shape and color (the basis of responding on Identity MTS).[8] These two bases of matching (same and shape/color) are pitted against one another in OMTSvRMTS such that the more likely adults were to infer matches on shape/color to be correct (that is what the incomplete object matches in OMTSvRMTS consisted of), the less likely, relatively speaking, they would be to match on the relation same—and vice versa.

Mechanisms of Change. It follows that the MTS tasks may have affected adults' preexisting inductive biases by one of two, not mutually exclusive, mechanisms: (1) *Inhibiting incomplete object matches*—a given MTS training task can increase the relative likelihood of inferring the relation same to be the correct bases of matching by making (incomplete) object, that is, shape/color, matches *less* likely to be inferred as correct bases of matching, or (2) *Promoting matches on the relation same*—the MTS training task can make inferring matches based on the relation same *more* likely. These mechanisms are not mutually exclusive. It is possible for a given training task to change inductive biases to decrease the likelihood of inferring shape/color object matches as the correct basis of matching, and for independent reasons also increase the likelihood of inferring same as the correct basis of matching. Either type of mechanism falls under Account 4.

Possible Effects of MTS Tasks in Experiment 3 on Adults' Preexisting Inductive Biases

Our task is to propose specific mechanisms, of the two types described above, through which Number and Size MTS increased the rate of matching on the basis of the relation same, whereas Color and Identity MTS had no such effect. The hypotheses as to the nature of these mechanisms detailed here are neither mutually exclusive nor exhaustive. Nevertheless they support empirical predictions, two of which are subsequently tested in Experiments 4 and 5. We believe all have merit and deserve empirical investigation.

Identity MTS

The near-ceiling rate of spontaneous success on Identity MTS suggests that this task is almost perfectly in line with the preexisting inductive biases of adults to match on shape and color. It follows that completing Identity MTS would not lead to any significant *changes* in inductive biases, leading to no effect on a subsequent OMTSvRMTS task.

---

[8] Identity MTS also involves a size match; however, given adults extremely low rates of spontaneous rates of success on Size MTS it is implausible that size matches drove the near-ceiling spontaneous success rate in Identity MTS.

Color MTS

In contrast to Identity MTS, a large proportion of adults did *not* spontaneously infer color to be the correct basis of matching in Color MTS and received a correction after making a mistake. This indicates that some adults initially attempted to match on a basis which was *not* color. There is only one object per card, and the figures are all approximately the same size (and adults are extremely unlikely to spontaneously match on size even when the size differences are large—only 10% did so on Size MTS). Therefore, it is very likely that the initial hypothesis for those who did not immediately match on color was a partial shape match (i.e., matching by some *similarity* in shape short of the two shapes being identical, a partial shape match—for instance both objects having right angles). Even some participants who succeeded on all trials may have initially looked for shape matches—but finding only partial ones switched to color as a basis of matching (that perfectly satisfied the logic of matching tasks), getting all eight trials correct.

Whatever adults' initial, incorrect, inference as to the correct basis of matching may have been, it is plausible that this basis of matching would be made *less* likely to be inferred as correct on a subsequent task, while color would be made *more* likely because it was the correct basis of matching (and those who made errors were explicitly told as much). Thus, if the large majority of those adults making errors on Color MTS initially inferred shape as the correct basis of matching the net effect of the task may have been to change their inductive biases so as to make them *less* likely to match on shape and *more* likely to match on color in a subsequent task. In OMTSvRMTS, however, incomplete object matches are *both* shape and color matches. As such, so long as Color MTS resulted in the inhibition of shape and promotion of color to equivalent extents the two effects would have no net effect on participants' likelihood of inferring incomplete object matches as correct in OMTSvRMTS.

Number MTS

Promoting Matches on the Relation Same. After being corrected on Number MTS, participants will have become *more* likely to match on the number of objects per card, since these were the explicit instructions provided. Given that all objects in Number MTS were identical *within* a card, there are two ways of interpreting matches on the number of objects per card being correct: (1), the number of objects on the sample card should match the number of objects on the choice card *or* (2), the number of *identical* objects on the sample card should match the number of *identical* objects on the choice card. On this latter hypothesis, we inadvertently made Number MTS trials progressive alignment trials for at least some participants. While the intended criterion of matching (number matches) is irrelevant to OMTSvRMTS (because all cards have two objects), conceiving the criterion in the second way would increase the likelihood of inferring relational matches as correct on OMTSvRMTS (because only relational matches have the same number of *identical* objects per card).

Another way that completing Number MTS might make adults more likely to infer that the relation same is the correct basis of matching on OMTSvRMTS might be through increasing the likelihood that a set property is the correct basis of matching in OMTSvRMTS. Sameness is a relation between individuals in a set, and not a property of an individual object.

Inhibiting Incomplete Object Matches. As with Color MTS, a significant proportion (41%) of adults did not succeed spontaneously on Number MTS. Given all objects in Number MTS were black, and the same size, it is implausible that adults' first, incorrect, hypothesis was that color or size was the correct basis of matching, given the logic of matching tasks. Rather, it is likely that those adults who did *not* spontaneously succeed on Number MTS (and potentially even some proportion of those who did succeed spontaneously) initially inferred *partial shape* matches to be correct. Receiving feedback *against* the possibility that shape matches are correct (i.e., being told they should match on number or seeing no perfect shape matches) may have made adults *less* likely to infer shape as a correct basis of matching on the subsequent OMTSvRMTS task. As shown by the baseline trials, sameness has the second highest likelihood of being inferred as the correct basis of matching, after shape/color identity matches, as the relevant basis of matching with stimuli such as these. Consequently, decreasing the likelihood that shape is the correct basis of matching will make sameness relatively more likely to be so.

Size MTS

Inhibiting Incomplete Object Matches. Size MTS training may have inhibited matches on shape and color in a similar way as proposed in the case of Number MTS: The overwhelming majority (90%) of adults did not succeed spontaneously on Size MTS. Indeed, 12% of participants made at least one further error even after being told the rule! Given that these participants were clearly not matching objects on size when they erred, and given the near-ceiling rates of spontaneous success on Identity MTS (that combines shape and color matches), it is highly likely that participants who did not succeed spontaneously initially inferred partial shape and/or color matches to be correct. After receiving instructions to match on size, participants may have become less likely to infer matches on shape or color as correct in a subsequent OMTSvRMTS task—leading them to become relatively *more* likely to infer matches on the relation same as correct.

Testing Specific Inductive Hypotheses

Our proposal is that identifying *specific inductive biases* and the *mechanisms by which they are changed* is not only important a priori to our understanding of relational reasoning but an empirically viable research program. To illustrate, we put our hypotheses regarding adults' preexisting inductive biases to a stronger test (Experiment 4) and test *one* of our hypotheses concerning one of the mechanisms through which MTS training might change preexisting inductive biases—namely, *that Number and Size MTS training had an inhibitory effect of on matching by shape and/or color in OMTSvRMTS (Experiment 5).*

Experiment 4

Method

Our hypothesis is that adults who do not match on relations in OMTSvRMTS have inductive biases that lead them to prefer matches on *shape and/or color specifically*. It follows that if we make shape and color less salient as matches in OMTSvRMTS, adults should become more likely to match on the relation same *even if* an incomplete object match—on shape and color—is available exactly as in the original task.

To decrease the likelihood that shape and color are inferred as the correct basis of matching, we can leverage the logic of matching tasks which stipulates that the correct basis of matching should be some feature that differentiates the choice cards (i.e., on which they are *not* equivalent), such that one choice matches the sample on that feature and the other does not. We can modify the OMTSvRMTS task to make all objects in each trial identical in shape and color, so that neither dimension comports with the logic of matching tasks. Of course, to retain the structure of OMTSvRMTS (i.e., an incomplete object match vs. a match on the relation same) objects must now vary on some less-salient dimension other than shape or color. An obvious candidate is *size*—a dimension that adults are extremely unlikely to spontaneously infer as the correct basis of matching (10% spontaneous success on Size MTS, Experiment 3; Figure 3).

It is important to highlight that this modified OMTSvRMTS task has *exactly the same structure and choices* as the task used in Experiments 2 and 3: The card with the incomplete object match has one object identical on all dimensions with the objects in the sample card while the other match exemplifies the relation same with objects that differ from both of those on the sample card. If adults' relational matching is determined by a *general* preference for object matches over relational matches (or vice versa) as suggested by previous authors (Simms & Richland, 2019; Vendetti et al., 2014), there should be no difference between performance on this modified OMTSvRMTS task and baseline performance (Experiment 2). In summary, Experiment 4 tests two interrelated hypotheses: (1) Adults' inductive biases are (at least in this case) specified at a more detailed level than a general preference for relations over object features and (2) The inductive biases of adults in our sample are *specifically* toward shape/color matches.

Materials

The modified OMTSvRMTS task was identical in format to OMTSvRMTS in Experiments 2 and 3, with the exception that color and shape were equated across all cards in each trial and objects varied only in size (see Figure 6). Specifically, figures throughout the task were black and on each trial all figures were the same shape. The sample card contained two identical relatively large figures, one choice card contained a relatively large figure (identical on *all* dimensions to the figures in the sample card; the incomplete object match) and one relatively small figure. The second choice card contained two identical relatively small figures (the relational match). The left-right position of the incomplete object match and relational match choice cards were counterbalanced across trials.

Participants

We recruited 204 participants from MTurk who had not participated in any RMTS study from our lab ($M_{\text{age}} = 36.56$, $SD = 11.87$).

Procedure

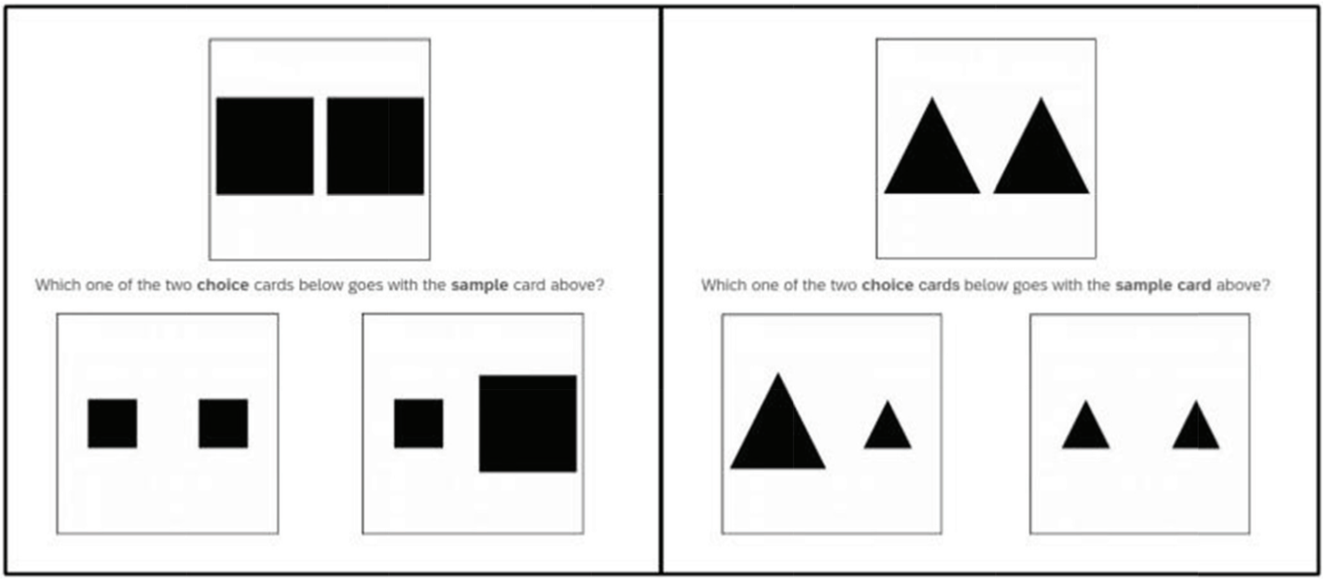
The procedure in Experiment 4 was identical to that of the baseline OMTSvRMTS task in Experiment 2.

Results

Removing shape and color as meaningful bases of matching nearly *doubled* the baseline likelihood that adults matched on the relation same as opposed to an incomplete object match (56% relational matches, Experiment 2, 93% relational matches in Experiment 4 (independent sample t test, $t(395) = 10.22$, $p < .0001$). This result supports both our hypotheses in showing that (1) Adults' inductive biases relevant to choosing between bases of matching in OMTSvRMTS were *not* formulated over object

Figure 6

Two Trials of the Modified Object Match-to-Sample Versus Relational Match-to-Sample (OMTSvRMTS) Task



Which one of the two **choice** cards below goes with the **sample** card above?

Which one of the two **choice cards** below goes with the **sample card** above?

matches or relational matches *in general*, but rather (2) over matches on shape and/or color vs. the relation same *specifically*. Once matches in shape and color, in particular, are made unlikely, adults overwhelmingly choose matches on the relation same over incomplete object matches. No account that assumes that preference for relational matches is established at the level of all relations vs. all object properties (e.g., Simms & Richland, 2019; Vendetti et al., 2014) can explain how the proportion of relational matches can go from 53% to 93% without the relational structure of the task changing *whatsoever*.

Experiment 5

Method

Experiment 4 provided evidence that those adults who did *not* make relational matches in OMTSvRMTS failed to do so as a result of inductive biases that lead them to infer shape and/or color as the correct bases of matching. This result is consistent with the possibility that one mechanism by which Number and Size MTS training increased relational matches in OMTSvRMTS was by changing adults' inductive biases such that they were less likely to infer color/shape as the correct bases of matching. The goal of Experiment 5 is to test the specific hypothesis that adults might have initially inferred partial shape/color matches as correct in Number/Size MTS then, upon receiving instructions that this was not the case, became less likely to infer shape and/or color—and more likely to infer the relation same—as being correct bases of matching in the subsequent OMTSvRMTS task.

A consequence of this hypothesis is that an MTS training task should be able to increase relational responding in OMTSvRMTS so long as (1) it has partial shape and/or color matches available (i.e., figures vary on shape and/or color) such that participants may *attempt* shape/color matches and (2) neither shape nor color is actually the correct basis of matching. Notice that on this hypothesis what the correct basis of matching in the MTS task actually *is* irrelevant; it just cannot be shape/and or color. This leads to a striking prediction: Adults should become more likely to engage in relational reasoning in OMTSvRMTS after completing an MTS task that has partial shape/color matches available (i.e., stimuli vary on shape and color) but has *no correct basis of matching at all*, that is, error feedback is randomly assigned for each trial. We call this a Random MTS task (see Figure 7). If the inhibitory mechanism we described for Number and Size MTS training was indeed part of the reason these tasks increased relational matching in adults, Random MTS should increase adults' relational matches in OMTSvRMTS by the very same mechanism. That is, we predict adults will initially infer shape and/or color as the correct basis of matching, then receive evidence that these are incorrect and hence become less likely to use these bases of matching in OMTSvRMTS, making them relatively more likely to match on relations.

Materials

Cards in each trial of Random MTS contained the same number of objects (one per card in seven trials, three per card in one trial). Objects on each trial were the same approximate size. On six trials all three objects were different colors, on two trials all objects were black. Objects on all trials differed in shape across cards. Which of the two choice cards was "correct" on any trial was randomized such that on half of the trials the left-side card was correct and on half of the trials the right-side card was correct.

Participants

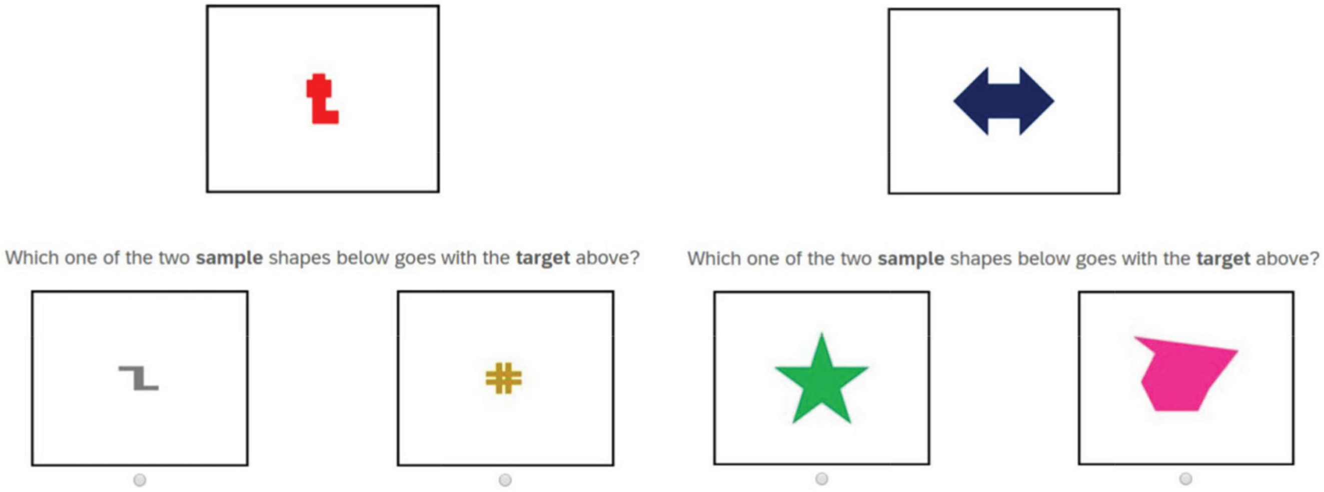
We recruited 196 participants via MTurk who had not participated in any of our MTurk studies on RMTS ($M_{\text{age}} = 36.68$, $SD = 11.36$).

Procedure

The procedure for Experiment 5 was identical to that of Experiment 3 except that participants completed eight trials of Random MTS as a training task. The subsequent OMTSvRMTS task was identical to that used in Experiments 2 and 3. Correction screens for Random MTS did not specify a correct basis of matching given there

Figure 7

Two Trials of Random Match-to-Sample (MTS)



Which one of the two **sample** shapes below goes with the **target** above?

Which one of the two **sample** shapes below goes with the **target** above?

Note. See the online article for the color version of this figure.

was no correct basis of matching in the task. If a participant chose what was randomly assigned as an "incorrect" choice card, they saw a correction screen which read "Good guess! But that was the wrong choice." No feedback was given after a "correct" choice.

Results

Training Task–Random MTS

Given the intentionally unsystematic nature of Random MTS, participants' scores on the task are not meaningful. The intended role of the task was to have participants choose *incorrectly* and receive error correction and/or to fail to find a consistent basis of partial shape/color matches across trials. In this, Random MTS was of course quite effective: Participants made an average of 4.91 "errors" (out of 8) and only one participant (out of 194) made no errors at all, while two participants made errors on all eight trials.

Test Task–OMTSvRMTS

An omnibus ANOVA across Experiments 2, 3, and 5, examined the effect of MTS training experience (none, Identity MTS, Color MTS, Number MTS, Size MTS, and Random MTS) on the percentage of relational choices on OMTSvRMTS (see Figure 5). There was a significant effect of training experience: $F(5, 1093) = 14.18$, $p < .0001$. Post hoc tests (Tukey's HSD) revealed that the proportion of relational matches in OMTSvRMTS did not differ across Baseline, Identity, or Color MTS training conditions. Likewise, the percentage of relational matches in OMTSvRMTS did not differ across Number, Size, and Random MTS training conditions—each of which elicited significantly more relational matches than did each of the Baseline, Identity, and Color MTS conditions. Thus, eight trials of Random MTS—with *no* systematic basis of matching—significantly increased the proportion of relational matches adults made on OMTSvRMTS compared with baseline (Experiment 2), confirming our hypothesis. Moreover the size of this increase is equivalent to the increases in relational matching as a result of training on Number and Size MTS.

An exploratory analysis tested for a relationship between the number of errors participants made in Random MTS (i.e., being told that their choice was incorrect) and the number of relational matches they subsequently made in OMTSvRMTS. These two variables were entirely unrelated $r(194) = .03$, $p = .68$. This suggests that either (a) a single correction inhibited shape/color as the correct basis of matching, with further corrections not meaningfully increasing this effect, or (b) that corrections were not strictly necessary; a task in which there were no consistent bases of partial color/shape matches that satisfied the logic of matching tasks across trials lead adults to rapidly infer that matches on shape/color are unlikely to be correct in the context of these tasks. Whether one of these hypotheses is correct, or both are, remains a question for future research.

The finding that eight trials of Random MTS (!) training leads to greater relational responding on a subsequent OMTSvRMTS task, relative to baseline, provides strong support for one of our hypotheses regarding the effects of Number and Size MTS. Specifically, we proposed that (at least part of) the reason Number and Size MTS training increased the likelihood of adults matching on the relation same in OMTSvRMTS is as a result of participants attempting partial shape/color matches on the former tasks and then receiving negative feedback which inhibits shape and color matches as correct in a subsequent task. Likewise, this result again confirms the hypothesis motivated by the results of Smirnova et al. (2015) and initially tested in Experiment 3—that relational responding can be increased by training tasks that do not involve second-order relational reasoning (do not involve matching same to same or different to different). Confirming such a dramatically counterintuitive prediction—that completing a task with no right answer, and that does not involve any relational matches, can increase spontaneous relational reasoning in adults—illustrates the value of specifying hypotheses at the level of specific, preexisting inductive biases and of exploring the mechanisms by which experience (such as MTS training) may change them.

Experiments 4 and 5 tests one class of mechanisms through which MTS training might change inductive biases to make relational responding on the basis of the relations same and different relatively more likely—by changing inductive biases so as to make otherwise salient shape and color matches less likely to be inferred as correct. Notice that the complementary kind of hypothesis, that is, making matches on same/different as more likely to be inferred as correct bases of matching, also leads to empirical predictions. For example, participants could have solved our Number MTS task using a rule like, "match cards which have the same number of identical objects on them" because all three-item arrays included identical objects. Doing so may have directly increased the likelihood relational matches were inferred as correct (i.e., directly promoted matching by "number of identical objects"). If this is the case, a Number MTS task where individuals within each card were *not* identical to one another should have a lesser effect on a subsequent OMTSvRMTS task.

General Discussion

The studies above contribute to the RMTS literature in three ways: **First**, they show that minimal interventions (*eight* trials of MTS training) drawn from comparative studies, where correct responses involved matches on the basis properties of arrays (number) or properties of individual figures (size) increased the likelihood that adults subsequently engaged in relational reasoning (matching on the basis of the relation same between elements in two different pairs of figures, Experiment 3). **Second**, the fact that only *some* MTS tasks increased second-order relational responding demonstrates that the representations involved in MTS do not include the relational representations required for RMTS. **Third**, two experiments explore the mechanisms by which Number and Size MTS training increased second-order relational responding, specifically the hypothesis that they inhibited preexisting biases to match on shape and/or color: *Eight* trials of Random MTS training, where there was no consistent rule to be found, and where participants were given random error feedback alone, increased relational responding to the same degree as Number and Size MTS (Experiment 5). Furthermore, simply equating shape and color across all objects in OMTSvRMTS increased relational responding from 56% (baseline, Experiment 2) to 93% (Experiment 4) despite the structure and choices (pitting incomplete objects matches against relational matches in a cross-mapping paradigm) being identical to the original task. We now review each of these conclusions in more detail, discuss how previous paradigms *may* have increased relational reasoning by changing inductive biases alone and how this latter possibility motivates further studies as to whether *population* differences in relational reasoning (e.g., human adults vs. Nonhuman animals, young children) may be due to inductive biases alone.

Changes in Inductive Biases Alone

Eight trials of training on Number, Size, or Random MTS increases the likelihood that adults engage in relational reasoning despite the fact that (1) these training tasks involved no second-order matches (in fact Random MTS involved no consistent basis of matching at all) and (2) adults *already had* the appropriate representations to succeed on RMTS—clear both from Experiment 1 and the fact that adults in this population have known the abstract meanings of the words "same" and "different" since age three (Hochmann et al., 2020). This leaves changing inductive biases as the *only* possible mechanism through which the training could have affected performance on the subsequent OMTSvRMTS task and provides a proof of concept that changing inductive biases *alone* can increase relational responding.

Representations of Sameness Underlying MTS Versus Those Underlying RMTS

Experiment 3 demonstrates that it is not the case that MTS tasks necessarily involve those abstract representations of *same* which are necessary for second-order relational matching. Neither Identity nor Color MTS training facilitated the use of such representations in a subsequent OMTSvRMTS test task (i.e., did not increase the proportion of relational matches adults made). This comports with the possibility proposed by Hochmann et al. (2016) and Zentall et al. (2018), discussed above, that sameness in MTS is realized by a match computation enacting a program like *store x*, *seek x* where *x* is a representation of the sample. In this procedure, abstractness is ensured by a lack of constraint on what entities can fill the variable *x*. The content *same* is implicit, in the sense of being carried by the match computation that underlies all acts of recognition and does not involve a mentally-represented symbol.

The mechanisms we propose for the effects of MTS tasks *do*, however, assume that participants represent not just the particular object (e.g., a blue square of a particular size and location within the sample card), but also have biases with respect to which of that object's features are relevant to the matching process (e.g., color, shape). At a minimum, participants must have a representation of at least one (or combination) of shape and/or color as a *class* of feature matches, such that completing Number, Size, or Random MTS can inhibit matching by this *class* of features. Experiments 3–5 strongly support the possibility of such an inhibitory process in adults (e.g., because the effects of Random RMTS are predicted only on such an account). Studies replicating the results of Experiments 3 (Kroupin & Carey, 2020) and 4 (Kroupin, 2020) with 4-year-olds (using RMTS and size-only RMTS tasks, respectively, as dependent variables instead of OMTSvRMTS) are consistent with this same inhibitory process in young children. Moreover, other evidence shows that nonhuman animals (here pigeons) can also learn to inhibit attention to a particular stimulus dimension when it varies systematically between trials but the discrimination rule being learned is over a different dimension (e.g., learning a rule based on color while learning to inhibit attention to pattern; Dopson et al., 2010).

Possible Mechanisms of Inductive Bias Change

As a result of Experiments 4 and 5 we have a good idea of at least one of the mechanisms by which Number and Size MTS changed adults' inductive biases to promote relational responding on a subsequent second-order reasoning task. Namely, tasks that either lead participants to attempt shape/color matches then receive negative feedback (Experiment 5) or make shape/color matches unlikely to be correct by equating stimuli on these dimensions (Experiment 4) dramatically increase adults' second-order relational responding in OMTSvRMTS. This is consistent with the hypothesis that Number and Size MTS had their effect precisely in the same way, that is, leading participants to attempt and then inhibit shape/color matches.

Possible Effects on Inductive Biases in Previous Paradigms

Looking back at previous training paradigms that have successfully bridged population differences in relational reasoning it is possible (though far from certain) that these too have had their effects by mechanisms changing inductive biases alone. For example progressive alignment (Kotovsky & Gentner, 1996) may have focused individuals on relational matches by having them initially co-occur with object-feature matches. Similarly, symbol training (e.g., Christie & Gentner, 2014; Premack, 1983; Thompson et al., 1997) may have made preexisting representations of sameness and difference more salient as bases of matching by mapping them to new symbols. Likewise, tens of thousands of trials of dogged training (e.g., Fagot & Thompson, 2011) may have gradually extinguished alternative hypotheses concerning the correct basis of matching, such that the animals finally arrived at preexisting representations *same* and *different* as hypotheses.

Smirnova et al. (2015) and Obozova et al. (2015)–Outstanding Questions

This brings us back to the original results of Smirnova et al. (2015) and Obozova et al. (2015): Does the fact that crows and parrots completed the same MTS tasks that increased relational responding in adults and children mean that, as in these latter populations, training lead to success by affecting certain inductive biases alone? No. For one thing, we cannot rule out that they may have succeeded without *any* training—we are not aware of any data regarding performance of these species on RMTS without training. Of course, untrained success would be an enormous outlier in the comparative literature. Likewise, it is possible that birds' success on RMTS was driven *entirely* by progressive alignment trials built into RMTS testing (i.e., where three out of four trials were reinforced progressive alignment trials, followed by one unreinforced RMTS trial) leading to generation of the requisite representations of sameness and difference de novo. That being said, the fact that birds performed approximately equally well on relational trials compared with progressive alignment trials from the very first session of testing weighs against a critical role for progressive alignment. Moreover, the role progressive alignment may have played is to change inductive biases *alone* by making preexisting representations of sameness and difference more salient for birds in the context of the task by having them constantly co-occur with object matches.

Thus, the issue of whether MTS training tasks (and/or progressive alignment) in Smirnova et al. (2015) and Obozova et al. (2015) was sufficient to produce success on RMTS in nonhuman animals by changing inductive biases *alone* remains an empirical question—one that the present studies certainly do not answer and which remains an important avenue for future research. Our work with human adults does, however, provide (a) a proof of concept that the kind of tasks used in these training paradigms *can* increase relational responding as a result of changing inductive biases alone and (b) an example of how we can generate and test hypotheses about the *mechanisms* by which this kind of training could have produced such an increase. Next we give a brief sketch of how we could go about such an analysis in the case of the training paradigm in the parrot and crow studies.

Smirnova et al. (2015) and Obozova et al. (2015)–Possible Effects of Training

How may the MTS training tasks which birds actually completed have brought about the necessary change in inductive biases? First, across the *multiple* MTS tasks birds will have learned that the correct basis of response could be *any one* of a number of possible features of the stimuli. This is a critical difference from other training studies in which nonhuman animals were initially trained on Identity MTS (e.g., Cook & Wasserman, 2007; Fagot et al., 2001). Specifically, learning *only* Identity MTS—where matches are made on color, shape and size—may lead animals to search for (partial) matches on these dimensions in a subsequent RMTS task, directing their attention *away from* relational matches and making RMTS success *more difficult*. Evidence from categorization-learning studies with nonhuman animals certainly supports the possibility that once subjects learn to categorize according to one dimension, attention to this dimension perseverates into subsequent tasks (Castro & Wasserman, 2016). If, as has been suggested to us, *most* nonhuman animals have experience with Identity MTS before participating in RMTS studies the attention to shape/color matches developed in the former may consistently interfere with performance on the latter. Needless to say, this is a critical issue to examine further and highlights the importance of explicitly detailing to subjects' previous training experience when reporting and interpreting RMTS performance.

Second, multiple MTS training tasks may have taught crows and parrots to strongly expect a perfect match between choice and sample on some dimension. This would likewise reduce the likelihood that birds would search for partial matches on some object feature. Third, training may have taught birds the logic of matching tasks—that is, that the correct basis of matching is one on which the correct choice card matches the sample and differs from the other choice card—once again focusing them on a search for perfect matches, which in RMTS occur only on the relations same and different. See Smirnova et al. (in press) for a convincing argument that this is at least part of the explanation for the success of their training regime.

These hypotheses are not meant to be an exhaustive list of possible aspects of the mechanism through which MTS training *may* have facilitated the birds' subsequent spontaneous success on RMTS. Rather, they are meant to illustrate how, using the kind of approach developed with adults above, we can analyze training paradigms that produce success on RMTS and develop testable hypotheses as to whether they may have had their effects through changes to inductive biases alone. Notice also that these hypotheses rely on inductive biases specified at the level of particular bases of matching—in contrast to the content-general bias toward relations suggested by Vendetti et al. (2014) and Simms and Richland (2019). We have demonstrated that, at least in the case of adults, the latter interpretation is implausible: No content-general bias would have resulted in a radical difference in percentage of relational matches (56% vs. 93%) between two tasks which are identical but for the dimensions on which the stimuli vary (i.e., OMTSvRMTS in Experiment 2 and 4, respectively). However, whether *population* differences in inductive biases—such as those bridged by the Smirnova et al. and Obozova et al. paradigms—are also differences at a specific level remains to be demonstrated elsewhere (see Kroupin, 2020; Kroupin & Carey, 2020, 2021).

Clearly, further research should explore what parts of the successful training regimes in the Smirnova et al. and Obozova et al. studies were necessary and/or sufficient for success at RMTS, and test specific hypotheses concerning the level of specificity at which and the mechanisms through which they affected subsequent relational responding. More generally, we see the current work as challenging the field, ourselves very much included, to specify in greater detail the mechanisms by which various training paradigms have their effects.

Though our results are consistent with the possibility of *some* population differences being due to inductive biases alone, we by no means wish to claim that profound differences in capacity do not exist s*omewhere* along both phylo- and ontogenetic spectra: Capacity accounts must be correct *in some cases*—after all, neither a nematode nor a neonate is likely to succeed on RMTS, regardless of how thoroughly we shift around their inductive biases. Likewise, sometimes population differences in relational reasoning depend upon the creation of new representations; the child cannot represent the relation "larger rational number" that holds between 1/2 and 1/4 until she has the concept *fraction*; a hard won achievement in both her learning history and the history of mathematics (e.g., in childhood, understanding of rational number is not achieved until between 8 and 12 years of age, after explicit instruction with a bootstrapping curriculum in school, see Carey, 2009, for review).

Rather, our point is that if differences can *sometimes* be the result of differences in inductive biases alone we must go through the process of identifying these preexisting inductive biases and testing whether changes to them are sufficient to produce success. Failing to do so may lead us to infer differences in capacities or representations where there are none.

Conclusion: The Importance of Inference

While attention to population differences in inductive biases is critical to the project of identifying the true origin-point of representational and computational capacities for relational reasoning in evolution and development, we wish to close by arguing against treating it *merely* as such. Knowing *when to use* the relational reasoning capacities and representations one has is just as integral to successful relational reasoning as is developing these capacities and representations in the first place. If we are interested in what is human-unique about relational reasoning, part of the answer will almost certainly lie in the *contexts in* and *readiness with which* we engage in relational reasoning. Neither of these are determined by our cognitive capacities or available representations per se, rather both processes guided by our inductive biases. Consequently, understanding how such inductive biases emerge over the course of ontogeny and phylogeny, as well as cultural history, is integral to the project of understanding human-unique relational reasoning. Given the scope and variety of inductive biases which individuals possess—be they humans or crows, adults or infants—the project of studying the structure of and changes to specific inductive biases may seem daunting. Yet, we propose that by choosing theoretically motivated case studies (such as RMTS here) we stand to make real progress on such issues. In closing, therefore, we wish to endorse the assessment of Michael Cole and his colleagues when faced with the not-unrelated challenge of exploring variation in cognitive capacities across cultural boundaries:

> [T]his is a cause for careful study, not despair.
>
> —(Cole et al., 1971, p. 22)

Context of Research

The approach of this work was shaped by Susan Carey's interest in the onto- and phylogenetic origins of abstract, combinatorial thought (in which the relations same and different play a central role) and Ivan Kroupin's interest in cross-cultural research that has shown that participants sometimes fail a task not because they cannot succeed, but because they make different inferences about that task than those presupposed by the experimenters. Exploring these issues in the case of RMTS adds to our understanding of how we can measure the true origin points of representations of sameness and difference across ontogeny and phylogeny (Susan Carey). Likewise, it provides a case study for research into the structured inductive biases resulting from evolution and/or (cultural) experience which can drive population differences in performance (Ivan Kroupin). Multiple follow-up papers are forthcoming, including work showing that the effects shown in Experiments 3 and 4 can likewise be used to lead four-year-old children—who ordinarily fail standard RMTS—to succeed on the task with no relational training, or even no training *at all*.

References

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Castro, L., & Wasserman, E. A. (2013). Humans deploy diverse strategies in learning same–different discrimination tasks. *Behavioural Processes*, *93*, 125–139. https://doi.org/10.1016/j.beproc.2012.09.015

Castro, L., & Wasserman, E. A. (2016). Attentional shifts in categorization learning: Perseveration but not learned irrelevance. *Behavioural Processes*, *123*, 63–73. https://doi.org/10.1016/j.beproc.2015.11.001

Christie, S., & Gentner, D. (2007). Relational similarity in identity relation: The role of language. In S. Vosniadou, & D. Kayser (Eds.), *Proceedings of the Second European Cognitive Science Conference*. Routledge.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373. https://doi.org/10.1080/15248371003700015

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*(2), 383–397. https://doi.org/10.1111/cogs.12099

Cole, M., Gay, J., Glick, I., & Sharp, D. W. (1971). *The cultural context of learning and thinking*. Basic Books.

Cook, R. G., & Wasserman, E. A. (2007). Learning and transfer of relational matching-to-sample by pigeons. *Psychonomic Bulletin & Review*, *14*(6), 1107–1114. https://doi.org/10.3758/bf03193099

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dopson, J. C., Esber, G. R., & Pearce, J. M. (2010). Differences in the associability of relevant and irrelevant stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(2), 258–267. https://doi.org/10.1037/a0016588

Fagot, J., & Thompson, R. K. R. (2011). Generalized relational matching by guinea baboons (*Papio papio*) in two-by-two-item analogy problems. *Psychological Science*, *22*(10), 1304–1309. https://doi.org/10.1177/0956797611422916

Fagot, J., Wasserman, E. A., & 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*(4), 316–328.

Ferry, A., Hespos, S., & Gentner, D. (2015). Prelinguistic relational concepts: Investigating analogical processing in infants. *Child Development*, *86*(5), 1386–1405. https://doi.org/10.1111/cdev.12381

Fodor, J. A. (1975). *The language of thought*. Crowell

Gentner, D., & Hoyos, C. (2017). Analogy & abstraction. *Topics in Cognitive Science*, *9*(3), 672–693. https://doi.org/10.1111/tops.12278

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355. https://doi.org/10.1016/0010-0285(80)90013-4

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature*, *410*(6831), 930–933.

Goodman, N. (1955). *Fact, fiction, and forecast*. Harvard University Press.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Erlbaum.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, *14*(11), 497–505. https://doi.org/10.1016/j.tics.2010.08.005

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action-based verb extension. *Child Development*, *82*(2), 674–686. https://doi.org/10.1111/j.1467-8624.2010.01567.x

Hochmann, J. R., Mody, S., & Carey, S. (2016). Infants' representations of same and different in match-and non-match-to-sample. *Cognitive Psychology*, *86*, 87–111. https://doi.org/10.1016/j.cogpsych.2016.01.005

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. https://doi.org/10.1016/j.cogpsych.2017.11.001

Hochmann, J.-R., Zhu, R., & Carey, S. (2020). *Does learning the words "same" and "different" require conceptual construction?* [Manuscript submitted for publication]. Department of Psychology, Harvard University.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. MIT Press.

Jamrozik, A., & Gentner, D. (2020). Relational labeling unlocks inert knowledge. *Cognition*, *196*, 104146–104146. https://doi.org/10.1016/j.cognition.2019.104146

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*(6), 2797–2822. https://doi.org/10.2307/1131753

Kroupin, I. (2020, January). *Not a lack of abstraction, but rather just distraction: Spontaneous success on a classic relational reasoning task in four-year-olds when shape/color information is removed* [Poster presentation]. Budapest CEU Conference on Cognitive Development, Budapest, Hungary.

Kroupin, I., & Carey, S. (2020). *You cannot find what you are not looking for: Population differences in relational reasoning are sometimes differences in inductive biases alone* [Manuscript submitted for publication]. Department of Psychology, Harvard University.

Kroupin, I., & Carey, S. (2021). Population differences in performance on Relational Match to Sample (RMTS) sometimes reflect differences in inductive biases alone. *Current Opinion in Behavioral Sciences*, *37*, 75–83.

Livins, K. A., & Doumas, L. A. (2015). Recognising relations: What can be learned from considering complexity. *Thinking & Reasoning*, *21*(3), 251–264. https://doi.org/10.1080/13546783.2014.954000

Long, B., & Konkle, T. (2017). A familiar-size Stroop Effect in the absence of basic-level recognition. *Cognition*, *168*, 234–242. https://doi.org/10.1016/j.cognition.2017.06.025

Long, B., Moher, M., Carey, S., & Konkle, T. (2019). Real-world size is automatically encoded in preschoolers' object representations. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(7), 863–876. https://doi.org/10.1037/xhp0000619

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J. N., Burkart, J. M., Call, J., Cantlon, J. F., Cheke, L. G., . . . Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), E2140–E2148. https://doi.org/10.1073/pnas.1323533111

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*(4), 431–467. https://doi.org/10.1006/cogp.1993.1011

Obozova, T., Smirnova, A., Zorina, Z., & Wasserman, E. (2015). Analogical reasoning in amazons. *Animal Cognition*, *18*(6), 1363–1371. https://doi.org/10.1007/s10071-015-0882-0

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*(2), 109–178. https://doi.org/10.1017/S0140525X08003543

Pepperberg, I. M. (1987). Acquisition of the same/different concept by an African Grey parrot (Psittacus erithacus): Learning with respect to categories of color, shape, and material. *Animal Learning & Behavior*, *15*(4), 423–432. https://doi.org/10.3758/BF03205051

Pepperberg, I. M. (in press). How do a pink plastic flamingo and a pink plastic elephant differ? Evidence for abstract representations of the relations same-different in a Grey parrot. *Current Opinion in Behavioral Sciences*.

Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences*, *6*(1), 125–136. https://doi.org/10.1017/S0140525X00015077

Rattermann, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In K. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 274–282). New Bulgarian University.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*(3), 249–271. https://doi.org/10.1016/j.jecp.2006.02.002

Shakespeare, W., & Furness, H. H. (1963). *As you like it*. Dover Publications.

Simms, N. K., & Richland, L. E. (2019). Generating relations elicits a relational mindset in children. *Cognitive Science*, *43*(10), e12795. https://doi.org/10.1111/cogs.12795

Smirnova, A., Obozova, T., Zorina, Z., & Wasserman, E. (in press). How do crows and parrots come to spontaneously perceive relations-between-relations? *Current Opinion in Behavioral Sciences*.

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology*, *25*(2), 256–260. https://doi.org/10.1016/j.cub.2014.11.063

Thompson, R. K., & Oden, D. L. (1995). A profound disparity revisited: Perception and judgment of abstract identity relations by chimpanzees, human infants, and monkeys. *Behavioural Processes*, *35*(1–3), 149–161. https://doi.org/10.1016/0376-6357(95)00048-8

Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching- to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 31–43.

Truppa, V., Mortari, E. P., Garofoli, D., Privitera, S., & Visalberghi, E. (2011). Same/different concept learning by capuchin monkeys in matching-to-sample tasks. *PLoS ONE*, *6*(8), e23809. https://doi.org/10.1371/journal.pone.0023809

Vendetti, M. S., Wu, A., & Holyoak, K. J. (2014). Far-out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological Science*, *25*(4), 928–926. https://doi.org/10.1177/0956797613518079

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, *25*(1), 161–169. https://doi.org/10.1177/0956797613502983

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*(1), 3–22. https://doi.org/10.1037/a0016327

Wasserman, E. A., Castro, L., & Fagot, J. (2017). Relational thinking in animals and humans: From percepts to concepts. In J. Call (Ed.), *American Psychological Association handbook of comparative cognition* (Vol. 2, pp. 359–384). American Psychological Association. https://doi.org/10.1037/0000012-017

Zentall, T. R., Andrews, D. M., & Case, J. P. (2018). Sameness may be a natural concept that does not require learning. *Psychological Science*, *29*(7), 1185–1189. https://doi.org/10.1177/0956797618758669

Appendix

Details of MTS Tasks in Experiment 3

Identity MTS

Stimuli: Each Identity MTS trial contained three cards—two bottom and one top. Each card contained one figure. The figure on one of the two bottom cards was identical on all dimensions with the top card. The figure on the other bottom card was of a different color and shape than the top card. All figures were the same height and width. All figures in the task were unique with the exception of identical figures on top and matching bottom cards on each trial for a total of 16 unique figures in the task.

The correct bottom card appeared on the left side of the screen on four trials and on the right side of the screen on four trials for a total of eight trials. If participants chose incorrectly, they received the message "Good guess! But that was the wrong choice. In this game, cards that have the same image go with each other." If participants chose correctly they received no feedback.

Participant age: *M* = 34.47, *SD* = 11.23

Color MTS

Stimuli: Each Color MTS trial contained three cards—two bottom and one top. Each card contained one figure. The figure on one of the two bottom cards was identical in color to the figure in the top card, but differed in shape. The figure on the other bottom card was of a different color and shape than the top card. All figures were the same height and width. All figures in the task were unique for a total of 24 unique figures in the task.

The correct bottom card appeared on the left side of the screen on four trials and on the right side of the screen on four trials for a total of eight trials. If participants chose incorrectly, they received the message "Good guess! But that was the wrong choice. In this game the cards that have the same color go with each other." If participants chose correctly they received no feedback.

Participant age: *M* = 35.57, *SD* = 11.61

Number MTS

Stimuli: Each Number MTS trial contained three cards—two bottom and one top. Each card contained either one figure or three figures. On cards with three figures all figures within the card were identical. The top card contained one figure on four trials and three figures on four trials for a total of eight trials. On each trial one of the two bottom cards contained three figures and the other contained one Figure All figures were unique in shape. All figures were the same height and width. All figures were the same color (black). All figures in the task were unique except those repeated within the same card for a total of 24 unique figures in the task.

The left/right position of the correct bottom cards was fully counterbalanced: On trials where the top card contained one figure, the correct bottom card (i.e., also displaying one figure) appeared on the left side of the screen on two trials and on the right side of the screen on two trials. On trials where the top card contained three figures, the correct bottom card (i.e., also displaying three figures) appeared on the left side of the screen on two trials and on the right side of the screen on two trials. If participants chose incorrectly, they received the message "Good guess! But that was the wrong choice. In this game cards with one image go with other cards that have one image and cards with three images go with other cards that have three images." If participants chose correctly they received no feedback.

Participant age: *M* = 35.54, *SD* = 10.98

Size MTS

Stimuli: Each Size MTS trial contained three cards—two bottom and one top. Each card contained one figure. The figures were one of two sizes—relatively big and relatively small, with the former roughly three times the height/width of the latter. The top card contained a relatively big card on four trials and a relatively small card on four trials for a total of eight trials. All figures were unique in shape and color for a total of 24 unique figures in the task.

The left/right position of the correct bottom cards was fully counterbalanced: On trials where the top card contained a relatively big figure, the correct bottom card (i.e., also containing a relatively big figure) appeared on the left side of the screen on two trials and on the right side of the screen on two trials. On trials where the top card contained a relatively small figure, the correct bottom card (i.e., also containing a relatively small figure) appeared on the left side of the screen on two trials and on the right side of the screen on two trials. If participants chose incorrectly, they received the message "Good guess! But that was the wrong choice. In this game cards with big shapes go with other cards that have big shapes and cards with small shapes go with other cards that have small shapes." If participants chose correctly they received no feedback.

Participant age: *M* = 35.72, *SD* = 11.21

Received July 22, 2019
Revision received February 11, 2021
Accepted February 18, 2021 ■