Inference in relational reasoning: A case study of Relational-Match-to-Sample

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Psychology

have examined a dissertation entitled

Inference in relational reasoning: A case study of Relational-Match-to-Sample

presented by Ivan Kroupin

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature Susan Elizabeth Carey

Typed name: Prof. Susan Carey

Signature Paul L. Harris

Typed name: Prof. Paul Harris

Signature

Typed name: Prof. Joseph Henrich

Signature

Typed name: Prof. Fiery Cushman

Signature

Typed name: Prof.

Date: April 29, 2021

Inference in relational reasoning: A case study of Relational-Match-to-Sample

A dissertation presented

by

Ivan Kroupin

to

The Department of Psychology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University

Cambridge, Massachusetts

April, 2021

Dissertation Advisor: Professor Susan Carey Ivan Kroupin

Inference in relational reasoning: A case study of Relational-Match-to-Sample

Abstract

Adult humans are uniquely proficient in reasoning with abstract relations (relational reasoning) - a capacity which underpins much of human-unique cognition, including scientific analogies, artistic metaphors and many phrases in day-to-day language and thought. An important question for cognitive science, therefore, is exploring the nature of this uniqueness - that is identifying the nature of population differences in relational reasoning between older children and adults, on the one hand and younger children and non-human animals on the other. The classic paradigm used to test for even the most rudimentary capacity for relational reasoning is Relational-Match-to-Sample (RMTS, Premack, 1983). RMTS involves matching pairs of objects on the basis of the relations same and different. As such, success on RMTS is evidence not only of a rudimentary capacity for relational reasoning, but also of representations of sameness and difference in unitary, symbolic format. If these representations are available early in life and/or across species, this would suggest that they may constitute primitives in an innate language of thought which underpins the combinatorial capacities of human language (and potentially non-human thought as well). Previous work has shown that non-human animals and children under the age of five struggle with RMTS, succeeding only after paradigms which may have led them to generate same/different representations *for the first time*. If this is the case, these populations are simply *not able* to engage in relational reasoning without special training. In contrast, a successful training paradigm used with crows and parrots provides a model of training (using simple Match-to-Sample, MTS tasks) which cannot have produced new relational representations. The effects of the original training were not, however, conclusive since other training mechanisms were also used with birds and these latter may have led them to generate same/different representations. The first paper demonstrates that MTS training can increase relational responding in human adults by changing *inductive biases alone* (since adults *already have* all of the necessary representations and capacities). This sets up the hypothesis that population differences in RMTS performance may sometimes *not* be differences in the availability in same/different representations, but rather differences in inductive biases *alone*, i.e. what bases of matching are inferred to be relevant - and that MTS training can bridge at least some population differences by changing these inductive biases *alone*. The second paper tests and confirms this hypothesis with four-year-old children, further predicting that children's failure on RMTS are in some cases due to inductive biases leading them to *specifically* seek partial shape and/or color matches. The third paper tests and confirms this final predictions by making shape and color unlikely bases of matching by holding them constant across pairs in RMTS, which results in *unprecedented wholly spontaneous success* by four-year-olds on the task. In sum, these results demonstrate that population differences in relational reasoning performance, and RMTS specifically, are sometimes differences in inductive biases alone. This implies that relational reasoning and same/different representations may be available earlier in ontogeny and phylogeny than previously thought - and highlights the crucial role of inductive biases in the development and deployment of relational reasoning capacities.

Table of Contents

Acknowledgments

No dissertation-writer is an island - and happily I am no exception. This work cannot have happened without the contributions of a great many people, all of whom I am grateful to and a few of which I will take a few lines to thank here.

First and foremost I am grateful to my advisor, Susan Carey. Her mentorship has been the key factor in my development from a slightly wild-eyed undergraduate with a tendency towards philosophizing and overlong theory sections into an altogether more competent graduate student able to express his ideas in clearer experiments and sharper prose.

Likewise, I owe thanks to my committee: Paul Harris, whose work has been a source of intellectual inspiration from the start of graduate school. Joe Henrich, for his support of my nascent cross-cultural interests over the past few years. And Fiery Cushman, for his readiness to engage with this work in a thoughtful and thorough way.

I would also like to thank the formative intellectual influences from my undergraduate years. Hugh Rabagliati and Alex Doumans, my advisors at the University of Edinburgh, were key figures in developing my interest in and allowing for my first experiences with empirical work. I am forever grateful for the confidence which they placed in me and my ideas early on - as well as for all of the support which allowed me to turn these ideas into meaningful research questions.

My undergraduate experience would also not be complete without the class on Marxist psychology taught by P Richard Shilcok. Many of the ideas discussed in that class remain important to me to this day and have found echoes in the work below. Finally I would like to thank David Levy for his class on Wittgenstein the profound influence of which has remained with me throughout my intellectual and personal life.

If no dissertation-writer is an island, neither are they a brain in a vat surviving only on nutrients and research. The support of peers and friends has been invaluable in keeping me in the reasonably happy and sane state of mind which allowed for the work described below.

All of my lab mates, past and present have been a great support and integral to my experience at Harvard, including Roman Feiman, Narges Ashfordi, Johnathan Kominsky, Bryan Leahy, Stephen Ferrigno and many others. I owe special thanks to Paul Haward, on whose friendship and intellectual companionship I've been able to depend from my first day in Boston.

I would also like to thank all of my close friends outside of academia, particularly the ST collective, for bringing much enough support and joy into my life throughout the process of graduate school. Without them the past six years would have passed in a smaller, grayer world.

Finally I would like to thank my parents, without whom none of this would be possible on a great many levels.

My dad for the personal and intellectual integrity which have been a model for me since before I can remember - and the deep curiosity which it is my joy to share with him to this day

My mom for her quiet strength and sense of purpose, which have been a guiding light through 28 years of thick and thin, and which I can only hope to have inherited in turn.

Thank you all.

“Someone is *imprisoned* in a room if the door is unlocked, opens inwards;
but it doesn't occur to him to *pull*, rather than push against it.”
Wittgenstein, 1977/1998, p. 48

“Why is a raven like a writing-desk?”
Carroll, 1865, p. 97

CHAPTER 1

Introduction

Background: The importance of relational reasoning

Relational reasoning

This dissertation concerns the nature of population differences in relational reasoning: The capacity to identify common structure among abstract relations which underpins analogy, metaphor and - by extension, much of human-unique cognition. For instance, Bohr famously posited a common structure between the relationships of planets to the sun in the solar system and of electrons to the nucleus in the atom. More generally, any time we apply mathematics to the real world (e.g. physics, statistics) we are comparing the interrelations between mathematical objects and the the interrelations of objects in the world (and not, e.g., digits to planets directly). Relational reasoning also pervades the lives of the non-mathematically inclined (or even non-numerate) in the form of metaphors and analogies baked into everyday language: We talk about life and love being a journeys, conflicts heating up, temporary solutions being like band-aids on a wound, avoiding unpleasant information as burying one's head in the sand - and so on. Given all of this, it is certainly fair to say that human cognition as we know it - and all of the many human civilizations and their products - would not exist without relational reasoning.

Not only is relational reasoning *necessary* to human cognition, it is also *unique* - at least in the degree to which it is evident in our achievements as a species. Human relational reasoning, like human culture, is *manifestly* more complex and productive than that of any other animals, "reflected in the fact that we're building cathedrals and walking on the moon, and they still sit naked in the rain." (Galef, in Grant, 2007, para 13, speaking of culture). Being the self-centered species that we are, identifying the nature of this unique capacity has become an important research program within cognitive science (e.g. Holyoak & Thagard, 1995; Halford, Wilson & Phillips, 2010; Kotovsky & Gentner, 1996, Premack, 1983).

An important angle of attack in this program has been to study *population differences in relational reasoning*. That is, human adults (or at least humans after a certain age) are the only population which manifestly engages in the kind of flexible, world-changing relational reasoning which motivates our interest in this subject. Consequently, by identifying what elements are necessary for this kind of reasoning other populations possess - and which they lack - we can identify not only what makes human relational reasoning as unique as it is, but also what changes must have occurred across phylogeny and regularly occur across ontogeny to produce adult human relational reasoning.

Relational-Match-to-Sample: A classic case study

A, if not *the*, central case study for investigating such population differences has been Relational Match to Sample (RMTS, Premack, 1983). The classic RMTS task (Figure 1) involves aligning two sets of the smallest possible size (two pairs) according to, arguably, the simplest of possible relations - sameness and difference (see e.g. James, 1890; Wasserman & Cook, 2010 on the importance of sameness to cognition).

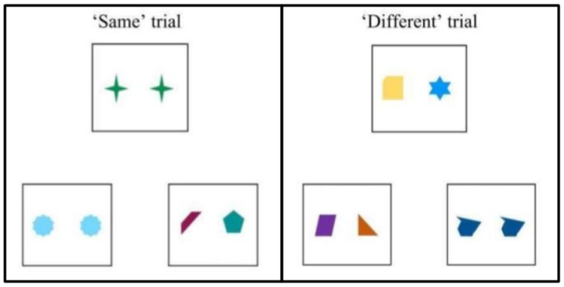


Figure 1: Examples of two RMTS trials.

At first glance, the pattern of populations which succeed spontaneously (i.e. with no training of any kind) on RMTS mimics the pattern of populations which engage in overt relational reasoning: Children above the age of five or so and human adults succeed spontaneously on the task (Hochmann et al., 2017 and Paper 1, below, respectively) while non-human animals and young children do not succeed spontaneously and perform above chance only with special, and sometimes very extensive training (see Papers 1,2 and Wasserman et al., 2017 for reviews). In sum, since the pattern of populations which spontaneously succeed on RMTS mimics, approximately, the pattern of populations that build cathedrals and walk on the moon (to say nothing of speaking in metaphors day-to-day), RMTS performance has face validity as an approach to investigating the nature of the kind of relational reasoning required for the latter. Moreover, inasmuch as RMTS involves, arguably, the simplest of possible relational comparisons, failures on the task have often been interpreted as an inability to engage in relational reasoning *at all* (e.g. Penn et al., 2008; Thompson, Oden & Boysen, 1997).

In addition to being an important measure of the emergence of relational reasoning across phylogeny and ontogeny, RMTS performance contributes to another, older literature. Specifically, since success on RMTS requires representations of sameness/difference in a format in which these representations can be compared with one another, failures on RMTS can be seen as evidence that such representations are *unavailable* in non-human animals and young children. If these representations are unavailable in these populations, this means that they are not candidates for innate, primitive constituents of a logic-like, language-like language of thought (Fodor, 1975).

This introduction has three parts. **Part I** reviews the relevant literature on the development of relational reasoning and primitive mental representations in reverse order. Prior to doing so it is worth acknowledging a possible objection up front: It may be immediately objected that the task lacks ecological validity for those populations which perform poorly. This is an important question which I will return to at length in **Part III** and in the **Conclusion** of the dissertation. For now, however, it is worth bracketing since the issue of ecological validity does not feature in the literature reviewed in the next two sections.

Part I

Population differences in representational and computational capacities

Language of thought

The older of the two strands of literature on which population differences in RMTS performance bear concerns the nature of innate mental representations. This question goes back in one form or another at least to Plato (with his notion of universal forms), and has consistently been part of the conversation about the nature and phylogeny/ontogeny of the human mind till this day (see e.g. Descartes, 1637/1985; Davidson, 1982; Bermudez, 2003; Darwin, 1871; Fodor, 1975). One major focus of this literature is the degree to which innate mental representations are in the form, exclusively, of sensorimotor primitives (as argued famously by Piaget, e.g. 1971), or whether these representations have some abstract content, such as ‘number’ or ‘agent’ (see Carey, 2009 for a detailed discussion). Fodor (1975; 2010), in particular, proposed that in order for the kind of combinatorial thought involved in reasoning and language to get off the ground there must be some set of innate abstract, primitive, representations which allow for other representations (e.g. of object, agents) to be combined in strings such as ‘these two trees have leaves of the same shape’ in a *language of thought*. Clearly, if some of these representations allow the combination of multiple elements, the former must be *relational* representations. Testing the hypothesis that there are innate, primitive relational representations involves empirically identifying *which* relations are innately represented in the mind and in *what format*.

Population differences in RMTS potentially play an important role in answering this question in the case of representations of sameness and difference - crucial elements in any combinatorial system (and certainly in the mind, see e.g. James, 1890). Specifically, the nature of the RMTS task means that success almost certainly depends up on representations of sameness/difference in the form of *abstract unitary symbols*, like the words "same" and "different", or some other symbol(s) with the same abstract meaning (e.g. *Ω,* as a symbol for same, and *λ,* as a symbol for different; see Papers 1 and 2 for this argument in full). Given such unitary symbols, the RMTS task requires matching only across individual representations - e.g. AA goes with BB or CD becomes *Ω* goes with *Ω* or *λ*. In effect, this transforms RMTS into a simple Match-to-Sample (MTS) task with correspondingly limited working memory requirements (see Thompson et al., 1997). MTS tasks, in turn, are within the capacities of even the simplest of animals (e.g. bees, Giurfa et al., 2001) and young infants (e.g. Hochmann, Mody and Carey, 2016).

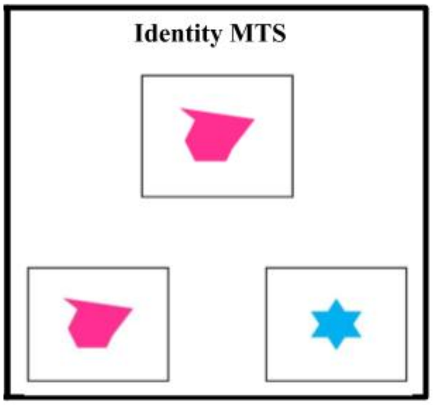


Figure 2: Example of an MTS trial drawn from Paper 2, in this case where the basis of matching is object identity (shape, color and size)

Sameness need not, however, be represented in the form of a unitary symbol. Hochmann et al., (2016), for instance, set out two alternative representational formats for sameness: First, sameness may be instantiated in a match computation with *no representation* of the relation at all. That is, an individual may solve an MTS task with the operation [store *X*, seek X] where *X* is a variable to be instantiated in a representation of the sample in a given trial (e.g., icon for the sample in Figure 2). There is no symbol for same in this procedure, only a symbol for the sample. The content same is carried in the match computation that underlies any act of recognition. This strategy is, in fact, how infants and pigeons solve MTS tasks (see Hochmann et al., 2016 and Zentall et al., 2018, respectively).

Another possible representational format for same-relations is a long-term memory representation of two specific, identical objects, [*X,X*]. For instance, a child may form a representation of two of their favorite spoons in a representation [*spoon, spoon*]. Subsequently any instance of sameness (within a pair) can be identified by aligning the objects in question with the internal representation of [*spoon, spoon*] (Hochmann et al., 2016).

Notice that while the unitary-symbol format of sameness is clearly sufficient for success on RMTS, the latter two formats are not: If the individual does not have a representation for the *outcome* of [store X, seek X], there is no sense in which this operation can be *iterated* as required for RMTS - i.e. first identify matches in one pair, then another pair, then identify the match between two successful match operations. If the individual *does* have a representation for the outcome of [store X, seek X], this *is* a unitary representation of sameness (as in Ω). The [X,X] format of representation does not have this problem, but would require an unfeasible number of working memory slots to solve RMTS since doing so involves holding in mind three pairs and the result of two comparisons (i.e. the long-term memory pair, sample pair, choice pair and the results of aligning the first two and the first and last).

In sum, populations which succeed on RMTS almost certainly have some form of unitary representation of sameness/difference, while those who fail may not. As such, population differences in RMTS may provide an important source of data for evaluating which formats of same/different representation are available to in the language of thought - i.e. innately, and in which species.

Emergence of relational reasoning

The more recent research program in which RMTS plays a role is one which is concerned with the nature and origins of relational reasoning. The modern origins of this work are intimately tied up with efforts to instantiate relational reasoning in computational models. Pioneers in the field such as Gentner, Forbus, Halford and Holyoak worked from the outset to implement the kind of relational reasoning described by their experiments in computational models (e.g. Holyoak & Thagard, 1989; Gentner & Forbus, 1991; Halford, 1993). Given the representational-computational framework such models provide, it is unsurprising that the population differences which have come out of this literature concern the relational *representations* (similarly to the language of thought issues above) and the *computational capacities* required to align them and draw inferences from this alignment.

Given its status as, plausibly, the simplest task which involves aligning sets according to (same/different relations) and not e.g. simply discriminating among different relations (see Paper 2 for a detailed discussion of such differences), success on RMTS is considered benchmark evidence of a population having the necessary representations and computational capacities to engage in relational reasoning. As a consequence, the particular computational/representational deficit which leads a given population to fail the task is indicative of the *kind* of difference which this population has with the relational reasoning of human adults (and older children). By extension, this contributes to our understanding of what *changes* are necessary across phylogeny and ontogeny to achieve the latter - and, conversely, what makes such relational reasoning unique. Specifically, the literature on the emergence of relational reasoning has posited three possible accounts of population differences in relational reasoning, of two *kinds* - capacity accounts and a learning experience account:

Accounts of population differences as differences in capacities

Account 1 - *Representational capacity*: Penn et al. (2008) proposed that failures v. successes on RMTS were a result of population differences in the capacity to generate representations of the necessary format. That is, according to this account, populations which do not succeed spontaneously on the task lack the capacity to generate stimulus-independent relational representations *altogether*.

Account 2 - *Computational capacity*: Halford (1993) proposed that population differences in relational reasoning performance were a function of working memory capacity. Richland et al., (2006) also proposed age-related differences as a function of increasing inhibitory capacities (i.e. children initially failed to match on relations because they could not inhibit attention to object features).

Note that both of these capacity limitation accounts of population failures assumed that no training training paradigms of any kind could bridge the differences between populations. As a result, successful training studies (e.g. Thompson, Oden & Boysen, 1997; Fagot & Thompson, 2011; Truppa et al., 2011; Smirnova et al., 2015; Obozova et al., 2015) rule out capacity limitation accounts of the previously observed failure of a given population. That being said it is almost certainly the case that many populations *in fact* do not have the necessary representational/computational capacities to succeed on RMTS (e.g. nematodes).

Accounts of population differences as differences in learning experience

Account 3 - *Generation of necessary representations*: Gentner and her colleagues argued in the context of both relational reasoning in general (e.g. Gentner, 1988) and RMTS in particular (e.g. Kotovsky & Gentner, 1996; Christie & Gentner, 2014; Ferry, Hespos & Gentner, 2015) that population differences were a result of differences in which groups had/had not had the appropriate experiences necessary to *actually generate* the relational representations necessary for success on the task. This account allows for training effects bridging population differences in RMTS performance insofar as training can produce the necessary representations of sameness/difference *de novo*.

Interim summary: What is at stake in account(s) of population differences in RMTS performance

If the three accounts above, alone or in combination, correctly characterize why all populations which fail RMTS (i.e. non-human animals, young children) do so, this rules out the possibility that unitary representations of sameness/difference are innate features of a logic of thought, for any species. Moreover, successful training studies rule out capacity accounts as explanations for *all* population differences in RMTS. It follows that if those populations which *can* succeed but require training initially fail because of a lack of the same/different representations in the necessary format and representations in this format are *unique* to older children and adults (and those individual animals/young children who have received the necessary training). Needless to say, such conclusions about the nature of innate primitive representations and human unique relational reasoning have important ramifications for psychological and philosophical issues stretching back centuries, if not millenia.

In fact, as discussed in Papers 1 and 2, the vast majority of training evidence is compatible with just such an interpretation (i.e. all population differences being a combination of the Accounts 1-3). A groundbreaking paradigm in the comparative literature, however, casts a shadow of doubt over this conclusion.

Part II

The birds' shadow: The Smirnova et al. (2015)/Oboova et al. (2015) paradigm

Smirnova et al. (2015) and Obozova et al. (2015) trained experimentally naive crows and parrots, respectively, using a complex paradigm with two types of training tasks: As a part of the RMTS testing procedure, this paradigm incorporates progressive alignment training. Progressive alignment is a paradigm pioneered by Kotovsky and Gentner (1996) in which relational matches co-occur in training with object matches (e.g. EE goes with EE or FG?), after which young children have been shown to succeed on relational matches alone (e.g. AA goes with BB or CD?, see Paper 2 for a more detailed discussion). Gentner and her colleagues (e.g. Kotovsky & Gentner, 1996; Gentner & Hoyos, 2017) propose that progressive alignment training allows the trainee to generate new 'relational abstractions'. As such, if birds succeeded due to progressive alignment alone this would be comfortably in line with Account 3 - i.e. the possibility that they *initially lacked* the representations of sameness/difference required for RMTS success, but *developed* these in the process of training.

Progressive alignment training was not, however, the only training birds received: In the first part of the paradigm, experimentally-naive birds completed a series of simple MTS tasks, learning to match first by color (black goes with black, white goes with white), then identity ('1' goes with '1', '2' goes with '2'), then number (cards with one shape go with cards with one shape, cards with three shapes go with cards with three shapes) and size (big shapes go with big shapes, small shapes go with small shapes) (see Smirnova et al., 2021 for details of this paradigm).

As a result of the full MTS and progressive alignment training paradigm, birds succeeded spontaneously (no differential reinforcement, that is, no error correction) on *three* separate RMTS test tasks - matching by same size, same shape and same color. Since progressive alignment was intermixed with RMTS testing, and birds succeeded from the first session of testing and maintained the same level of performance in subsequent sessions, any effects of progressive alignment must have occurred *extremely* rapidly. Were this the case, it would present by far the smallest amount of relational training necessary for success in any comparative study.

At first glance, this success may appear to be attributable to the remarkable efficiency of progressive alignment as a method for producing new relational representations. A closer look at the data, however, casts doubt on this interpretation. Specifically, *from the very first session* both crows and parrots succeeded on progressive alignment and RMTS test trials *at equal rates* (83%/76% and 72%/75% average performance on RMTS/progressive alignment for crows and parrots, respectively). This suggests that either progressive alignment *almost immediately* produced relational representations which had heretofore not existed in these species' cognitive repertoires, or that some other aspect of the training paradigm facilitated their rapid RMTS success.

As discussed above, the other part of the Smirnova et al. (2015)/Obozova et al. (2015) training paradigm were a series of simple MTS tasks. This training, however, differs qualitatively from all other forms of training used in RMTS experiments since it *does not involve the use of same/different representations*. As discussed above, Hochmann et al., (2016) proposed that MTS tasks can be solved using a procedure like [store *X*, seek *X*, where the variable *X* stands for a sample that can be encoded. Moreover, the same authors showed that this is *in fact* the way that infants solve MTS - and Zentall et al., (2018) demonstrated that pigeons do so as well.

Thus, the results of Smirnova et al./Obozova et al. (2015) throw a wrench into the established set of accounts three accounts which explained population differences in RMTS - and the corresponding conclusions about the nature of innate mental representations and human unique relational reasoning: Needless to say, both capacity accounts were ruled out by the mere fact of birds' success. The representational-learning account, previously capable of account for all training effects, now faced the unhappy choice of assuming either 1) That progressive alignment could produce qualitatively new representations in birds *de novo* after just a few trials, or 2) Conceding that MTS training tasks were critical to birds' success, and that training tasks in some cases could lead to success *without producing new relational representations*[1]. If the latter is the case, we must assume that such representations *already exist* in at least some populations which ordinarily fail RMTS - and that the three existing accounts of population differences can no longer explain the full data pattern of RMTS performance across populations. Given the implausibility of just a few trials of progressive alignment producing qualitatively new representations of sameness and difference in non-human animals, this means that we require a new account of population differences, and a corresponding account of how training (MTS training in particular) can overcome such differences.

Part III

Searching for a different kind of difference

A detour into cross-cultural psychology

The work discussed thus far has considered differences across species, across development and across learning experiences within Western, educated samples. Given that the kind of difference needed to explain the Smirnova et al., (2015)/Obozova et al. (2015) results had

[1] There remains, technically, also the possibility that MTS training produced new relational representations in birds *despite the fact* that Zentall et al. (2018) demonstrated that success on such tasks in non-human animals did not *require* such representations. This possibility is reviewed in detail in the discussion of Paper 2.

not emerged from this work, it makes sense to look for a different kind of difference along another dimension of contrast altogether. In this case, the dimension of contrast I propose to be most useful is in *cultural background*, and more specifically in experience with formal education. The literature regarding such differences is enormous (see Rogoff, 1981; Cole, 1990 for reviews), but one example will be sufficient to serve as an illustration of a different *kind* of difference prevalent in this work.

Working in Uzbekistan during the early Soviet Union, Alexander Luria (1976) ran a series of studies attempting to identify the effects of the rapid industrialization, including the introduction of formal schooling, on the cognition of the local population. Among the striking effects he found was that non-schooled participants persistently failed to draw logical conclusions from syllogisms with unfamiliar premises. For instance, when presented with the syllogism "In the North where there is snow all bears are white. Novaya Zemlya is in the North. What color are the bears there?" participants without formal schooling would persist in responding from personal experience (e.g. "I don't know, I've never been there") as opposed to drawing the conclusion motivated by the premises. An analogous effect was found with young children, who failed to draw logical conclusions from syllogisms whose premises contradicted their own experience (e.g. "All cats bark. Rex is a cat. Does rex bark?", Harris, 2001).

Subsequent work with both non-schooled populations (Diaz, Roazzi & Harris, 2005) and children (Harris, 2001), demonstrated that the source of these 'errors' in both populations was *not* they result of an *inability* to make inferences (*contra* Luria's initial conclusions) but rather differences in *biases* as to how the question-context was interpreted. That is, non-schooled groups and young children in Western samples were apt to adopt what Scribner (1977) referred to as an "empirical bias" - in effect an assumption that the *point* of answering a question was to bring to bear one's own world knowledge. From this perspective, the logical structure of the premises is simply an odd way of talking, irrelevant to the point of the conversation (which is *about* e.g. relevant bear-knowledge). Evidence for this comes from the fact that the majority of non-schooled participants did not reproduce the logical structure of the premises when simply asked to repeat them back to the experimenter (Scribner, 1977).

Harris and his colleagues (Harris, 2001; Diaz, Roazzi & Haris, 2005) were able to dramatically increase the rate of logical responses in both non-schooled groups and young children - that is bridge the population difference between them and schooled adults - simply by prefacing the syllogism with a phrase intended to inhibit the use of the participant's own experience (e.g. "Imagine a planet where…"). In other words, the original population difference was *not* one of capacity, *nor* was it related to what representations were available to participants at the time of testing. Rather, the difference which the experience-inhibiting preface addressed was one in *inductive biases* - a set of cognitive structures designed to guide inferences as to what capacities and representations should be brought to bear on the current context. Clearly, in addition to differences in capacities and available representations, population differences in task performance - even differences between spontaneous success and outright, persistent failure - can be a result of a *difference in inductive biases*. This is certainly a different kind of difference than the ones discussed in the literature reviewed thus far since it involves *neither* differences in representations *nor* in the computational capacities required to manipulate them - and, in fact, can occur in cases where the latter two are held constant (as is the case in the work of Harris and colleagues).

Inductive biases in relational reasoning

If inductive biases are to be a plausible source of population differences in relational reasoning, we must establish that they play a role in the process of such reasoning in the first place. Fortunately, this is easily done. Given even a small vocabulary of relational representations, there is an infinite number of possible specific relational comparisons one can notice in any stimulus. Given, say, a raven and a writing desk, one can attend to an infinity of relational features that the two have in common - e.g. both being larger than a snail, both being less dense than a neutron star, both being less likely to win a Grammy than Ray Charles, Bob Dylan or Billie Eilish, etc. *ad inf.* Identifying *which* of such an infinity of relations is the relevant one to attend to in context consequently poses an enormous inductive challenge - compounded by the fact that there are *also* an infinity of non-relational features ravens and writing desks have in common (e.g. containing carbon atoms, having no portholes or, as suggested by Carroll, being able to produce a few notes). Separating out which predicates are relevant in context is a problem of the same kind that Goodman (1955) famously discussed in this green/grue example.

According to Goodman, there is no way to know, looking at an emerald, that the appropriate predicate to apply to it is "green" and not "grue" (where "grue" means green until time *t*, and blue afterwards). There are, moreover, an infinity of such predicates in any given case. The only way we decide which predicates to apply, Goodman concludes, is through a process of *entrenchment* - some predicates become more likely to be used because they have previously been usefully deployed in similar contexts. Goodman's problem applies directly to relational reasoning since any problem that applies to predicates in general *a fortiori* applies to *relational* predicates. Returning to RMTS, there is no way for us to know *a priori* that the correct basis of matching is sameness and not, say, whether the shapes on each card have greater or fewer than five edges. In other words, answering the question in RMTS - "Which of these two [choice] cards goes with this [sample] card?" is no more constrained *a priori* than, say, "Why is a raven like a writing desk?" (Carroll, 1865, p. 97). Consequently *spontaneous* success on RMTS is a result of having the right constellation of *entrenched inductive biases*.

In sum, relational reasoning has *at its foundation* an inductive problem - as Livins and Doumas (2015) perceptively put it "one must recognise [i.e. infer as relevant] a relation before one can map it to another relation, and so without recognition, the rest of the [relational reasoning] process would not even get off the ground." (p. 252). It follows that, at least in principle, inductive biases may differ across populations such that one population spontaneously succeeds on a task like RMTS, while the other fails while *holding constant* their capacities and available representations (contra Accounts 1-3).

Relevant work within the relational reasoning literature

Several studies in the recent literature on relational reasoning have appealed to inductive biases as a source of difference in relational reasoning performance - albeit not in these terms. One approach demonstrated that relational responding can be increased in either as a result of having participants complete analogies prior to the task (e.g. wood:woodstove::furnace:?) (Vendetti et al., 2014, with adults; Simms & Richland, 2019, with four-year-olds). These authors described group differences in terms of degrees of overall relational 'mindset' - in our terms a domain-general inductive bias leading individuals to infer relational comparisons as relevant (over object comparisons).

Another approach aimed to increase young children's relational responding by modifying RMTS tasks so as to place relational responses in a causal context (Walker & Gopnik, 2014; Walker, Bridgers & Gopnik, 2016; Carstensen et al., 2019; Goddu et al., 2020). These proposals also imply that population differences in relational reasoning are a function of differences in inductive biases - such that young children's inductive biases do *not* lead them to infer relational responses as correct in the standard RMTS task but *do* so in causal variants.

Paper 3 provides a detailed discussion of the various inductive bias accounts in the literature (including the one proposed in Papers 1 and 2). For the moment, however, it is enough to note that both cross-cultural evidence from other lines of research and emerging work in the relational reasoning literature itself motivates a *fourth account of population differences*.

A different kind of difference - identified

Specifically, we can now formulate a new kind of population difference in addition to accounts positing differences in representational or computational capacity and the availability of specific representations. This new account falls, like the specific-representation account - under the umbrella of <u>Accounts of population differences as differences in learning experiences</u>:

Account 4 - *Differences in inductive biases alone*

Population differences in RMTS performance may be, at least in some cases, differences in inductive biases *alone*. That is, humans over the age of five in Western populations clearly have inductive biases which would lead them to spontaneously infer sameness and difference as the correct bases of matching in RMTS, as evidenced by their spontaneous success on the task. Other populations, however, may have different inductive biases and thus fail to infer sameness and difference as the correct bases of matching *despite being perfectly capable of success on the task*. This is intended as a direct parallel to the *kind* of failure non-schooled populations and young children display when presented with unfamiliar/counterfactual syllogisms. As in the syllogism case, the proposal here is that a training paradigm may produce success on RMTS in a population which ordinarily fails *without* changing capacities or producing new relational representations - by changing the relevant inductive biases *alone*.

Part V

The present work: Smirnova et al./Obozova et al. (2015) redux

We now return to the perplexing results of Smirnova et al. (2015) and Obozova et al. (2015) in which simple MTS training apparently played a role in facilitating spontaneous relational responding in non-human animals. Since completing MTS *does not require* relational representations (Hochmann et al., 2016; Zentall et al., 2018) these tasks almost certainly cannot lead birds to generate representations of sameness/difference *de novo* - much less introduced new representational/computational capacities (see Papers 1 and 2 for this argument in detail).

There is no in-principle reason, however, that these tasks could not have produced success in birds by adjusting their pre-existing inductive biases such that they became more likely to infer sameness/difference as the relevant bases of matching. As we can see, identifying *whether* and *how* these MTS tasks can, in fact, change inductive biases so as to increase relational responding in RMTS - especially in populations which otherwise fail the task - presents itself as a crucial test not only of a possible mechanism underlying the success of birds in Smirnova et al/Obozova et al., but of the general hypothesis that population differences in relational reasoning can be the result of inductive biases *alone*. Moreover, assessing the effects of each MTS task used in the bird studies *individually* allows us to isolate the *mechanisms by which* they increased relational responding (or failed to do so).

In sum, an exploration of the effects of MTS training tasks on relational matching presents itself as a critical opportunity to explore the possibility that population differences in relational reasoning are sometimes due to inductive biases alone. This would rule out the possibility that relational reasoning and unitary, symbolic representations of sameness and difference are *unique* to humans over the age of five, and once again open the question of whether such representations are a part of an innate language of thought.

Paper summary

All three papers focus on exploring the case study of Smirnova et al. (2015) via the lens of population differences in, and training effects on inductive biases. Specifically, the work explores the possibility that MTS training tasks lead to changes in inductive biases so as to make relational matches more likely to be inferred as correct and, by extension, that initial failures in certain populations are due to predictable inductive biases which can be changed with target interventions.

Paper 1 tests the hypothesis that certain MTS training tasks from Smirnova et al. (2015) can increase the likelihood that *human adults* make relational matches. Given this population is clearly *capable* of making such matches (as confirmed in Experiment 1 of Paper 1), any effects of MTS tasks must take the form of adjusting adults' inductive biases. Training adults on only one of the MTS training tasks from Smirnova et al. (2015) allows us to test not only *whether* these tasks change inductive biases so as to make relational matching more likely, but also identify *which* of these tasks do so.

Furthermore, the proportion of adults who spontaneously succeed (all trials correct, no correction received) on a given MTS task gives us an indication of how much this populations' pre-existing inductive biases align with the particular basis of matching (e.g. size of geometric shapes in Size MTS). This, in turn, allows us to test specific hypotheses as to *how* MTS training tasks may change these pre-existing biases so as to increase relational responding in adults.

Paper 2 trains four-year-old children, a population which does not spontaneously succeed on RMTS, on one of three MTS tasks drawn from Smirnova et al. (2015) prior to testing on RMTS. This allows us to test two hypotheses: A) Population differences are sometimes differences in inductive biases *alone* - in this case four-year-olds initially fail RMTS because of differences in inductive biases from older children/adults and B) that the mechanisms by which four-year-olds inductive biases can be changed so as to increase the likelihood of inferring relations to be the correct bases of matching are at least somewhat continuous with the equivalent mechanisms in adults. As with Paper 1, the proportion of children spontaneously succeeding on a given MTS task will also give us information regarding four-year-olds' pre-existing inductive biases.

Paper 3 begins with the information gained in Paper 2 about four-year-olds pre-existing inductive biases in matching tasks - namely that their biases lead them to infer shape and color as the correct bases of matching. This warrants the hypothesis that children at this age may fail RMTS as a result of attending to color and shape and *not* focusing the relations same and different within cards. This hypothesis is tested by designing a modified RMTS task in which shape and color matches are either impossible or uninformative, with the prediction that children should be much more likely to engage in relational reasoning once the *specific* dimensions made salient by their pre-existing inductive biases are removed.

Despite this work concerning only one task, the paradigmatic nature of RMTS means that the stakes of this research program are significant: If, at least some cases, population differences in RMTS performance can be shown to be differences in inductive biases *alone*, such data will 1) mean we cannot rule out (on the basis of RMTS) the possibility that same/different representations are primitive elements of an innate language of thought 2) demonstrate that the combination of representational format and computational capacities required for relational reasoning is *not* unique to humans over the age of five and 3) illustrate the vital importance of studying the nature of inductive biases involved in relational reasoning.

CHAPTER 2

Inference in Relational Reasoning: Relational Matching in Adults as a Case Study

Introduction

"All the world's a stage, And all the men and women merely players"

(Shakespeare, trans. 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, Wilson & Phillips, 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, Castro & Fagot, 2017), i.e., 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 non-human 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 second-order relational reasoning, as it requires computing relations between relations - i.e. whether two pairs both instantiate the relation same or both instantiate the relation different, in spite of not sharing any individual objects in common.

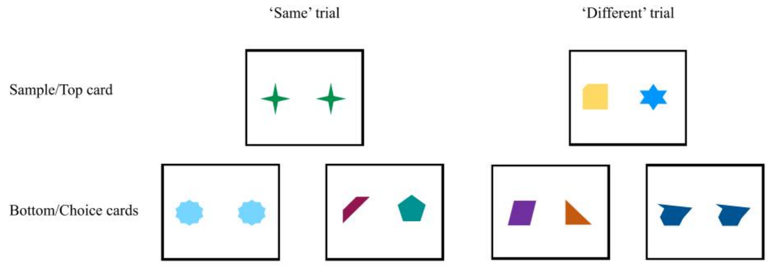


Figure 1: Examples of two RMTS trials.

While the matches involved in Premack's RMTS may appear simple (and are, in fact, extremely easy for human adults in a US sample - see Experiment 1), success on the task has proven to be a tremendous challenge for non-human animals (see Wasserman, Castro & Fagot, 2017 for a review) and children under the age of about five (see Premack, 1983; Hochmann et al., 2017; Kroupin & Carey, under review for evidence of RMTS failures at four years[2]). 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 ten months (Hochmann, Mody & Carey, 2016) relatively quickly learn MTS tasks (i.e. A goes with B or A). That is, these

[2] 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 (under review) for a discussion of RMTS variants.

populations understand matching tasks in general, but many more complex animals and much older children fail to match on the basis of relations, i.e. 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 (Premack, 1983; Hochmann et al., 2016; Zentall, Andrews & Case, 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 ten-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 US population, Hochmann et al., under review), one abstract representation of sameness is literally *same*, i.e. a mental representation of the word "same" in the mental lexicon. Another example is the mental representation of a heart-shaped figure which chimps were taught to recognize as mapping to pairs of identical objects (Thomspon, Oden & Boysen, 1997).

In sum, 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 ten-month-old humans, the latter is *at best* a serious challenge even for primates and human children as old as five (Premack, 1983; Penn et al., 2008; Thompson & Oden, 1995; Wasserman et al, 2017; Hochmann 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 Paper

The current paper 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, Holyoak and Povinelli (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 non-human 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 non-human 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 sum, Penn et al.’s hypothesis is that non-human 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 - e.g. vertebrates, including non-human 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 (for example, of attention to non-relational properties of stimuli, e.g. Richland, Morrison & Holyoak, 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; e.g., 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*[3]. 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

[3] In practice, the format of available representation (Accounts 1,3) 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 Thomspon et al., 1997; Halford, 1993). 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.

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 Account 1 and Account 3 is that RMTS failure on Account 1 result from a lack the capacity to create relational representations which 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 v. 3 distinction with a different example: Both a dung beetle and a typical, up-and-coming nine-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*, i.e. 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* which 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, e.g. 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 non-human animals. We believe that this literature rules out Accounts 1 and 2 (capacity limitation accounts of failures) for at least some populations which 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 non-human 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 thus 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 thus 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, Oden & Boysen, 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 three- to four-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, i.e., "material." This was true even though he did not know the words "plastic," pink," "brown," "elephant," or "giraffe" (Pepperberg, 1987; 2020). 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 non-human 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 twenty-nine individuals) performed above chance on RMTS (Fagot & Thompson, 2011; see also Truppa, Piano Mortari, Garofoli, Privitera, & Visalberghi, 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 non-relational 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 pre-existing 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 non-relational strategies.

Gentner and her colleagues propose that progressive alignment produces new relational representations (Account 3) - e.g. “analogical comparison [including progressive alignment] is... the main driver of *new relational abstractions.*” (Genter & Hoyos, 2017, p. 687, emphasis added, see also Kotovsky & Gentner, 1996; Ferry, Hespos & Gentner, 2015). 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, Imai & Okada, 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 pre-existing 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 Match-to-Sample (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., (2020), 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 non-differentially reinforced test trial in which objects matched only on approximate size (e.g. x goes with y or Z). Birds succeeded on the non-differentially 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 non-differentially 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 thus 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 - i.e. that differences between untrained crows and parrots, on the one hand, and human adults, on the other is in the absence v. 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 to 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, under review, *most if not all*) other comparative studies included MTS training prior to test, these did *not* facilitate the use of relations in a subsequent matching task (e.g. MTS-trained baboons in Fagot, Wasserman & Young, 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 Hochman et al., Zentall et al.), 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, Wasserman & Young, 2001). In sum, 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 non-relational 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 (which 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

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 which 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 in order 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 (mean age = 34.65, SD = 10.36). Participants were recruited from the US 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, and all participants gave informed consent prior to 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 for examples 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?”[4] .

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.

[4] 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, i.e. those from among which the match is *selected*. In order 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.

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 7 of 8 relational responses (statistically above-chance performance, binomial test, p = .04). Since there was no error feedback on this task, this means that almost all (93%) individually succeeded above chance, i.e., 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 which 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 etc. - and were located in various positions on the displays. In contrast, the figures in our study were colored geometric shapes which 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 & Wassserman, 2013).

In sum, 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 which involves comparisons between sets of colored geometric shapes.

Experiment 2

Adults' ceiling-level performance on RMTS leaves little room to change their pre-existing 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 Ratterman & 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, Wu and Holyoak (2014) have shown that adults' choices were evenly split (50%) in a task which 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[5]. 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 (which 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 non-relational basis of matching *over* a relational one. Specifically, our new task ("Object Match-to-sample v. Relational Match-to-Sample", OMTSvRMTS) is identical to that of Christie and Gentner, except that the non-relational 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* - i.e. an incomplete object match - and the other differed from

[5] 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*.

the objects on the top card both in shape and color (see Figure 2, and Appendix for full description of the stimuli).

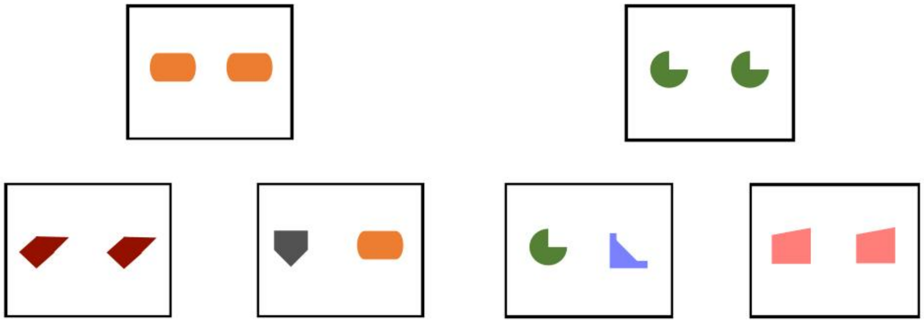


Figure 2. Two OMTSvRMTS trials. In the left triad, the relational match is the card on the left and the incomplete object match is the card on the right. In the triad on the right, the incomplete object match is the card on the left and the relational match is the card on the right.

Participants

We recruited 193 adult participants from Amazon Mechanical Turk who had not participated in any RMTS study from our lab (mean 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[6].

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

[6] 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. In order 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”.

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% 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 sum, 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

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 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.[7] There was no feedback on the subsequent OMTSvRMTS test task.

[7] 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, under review). 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.

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, prior to 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 pre-existing 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% (Figure 4). Combining these data regarding adults' pattern of pre-existing 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

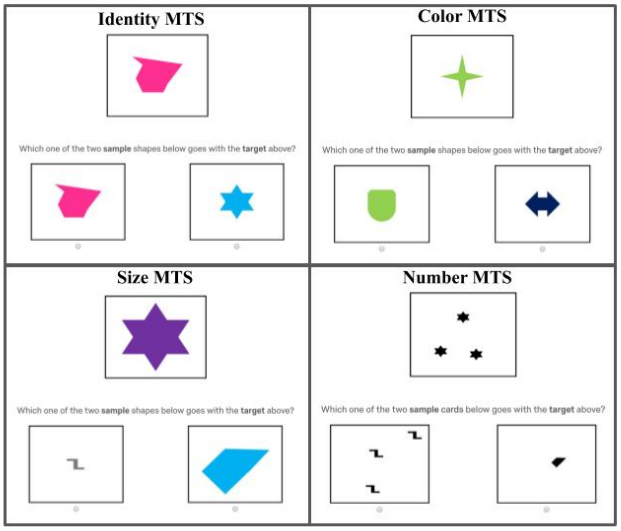
Participants were recruited via MTurk as in Experiments 1 and 2. None had participated in any previous (OMTSv)RMTS study in our lab. Each of Experiments 3A-D had two sample sizes. First, the total number of participants who completed the task was used to analyze spontaneous success rates on MTS tasks. Second, since 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 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, i.e. 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 3) 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, 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., 2021, 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.



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?

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

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

Figure 3. Example trials of MTS tasks in Experiments 3A-D.

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 4. The proportion succeeding spontaneously on RMTS in Experiment 1 is also included for comparison. Note that for RMTS in Experiment 1, since 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 in these analyses. The proportion of spontaneous succeeders differed across tasks $\chi^2(4, N = 1352) = 426.57$, $p < .0001$. 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' pre-existing 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 pre-existing 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 succeders 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' pre-existing 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.

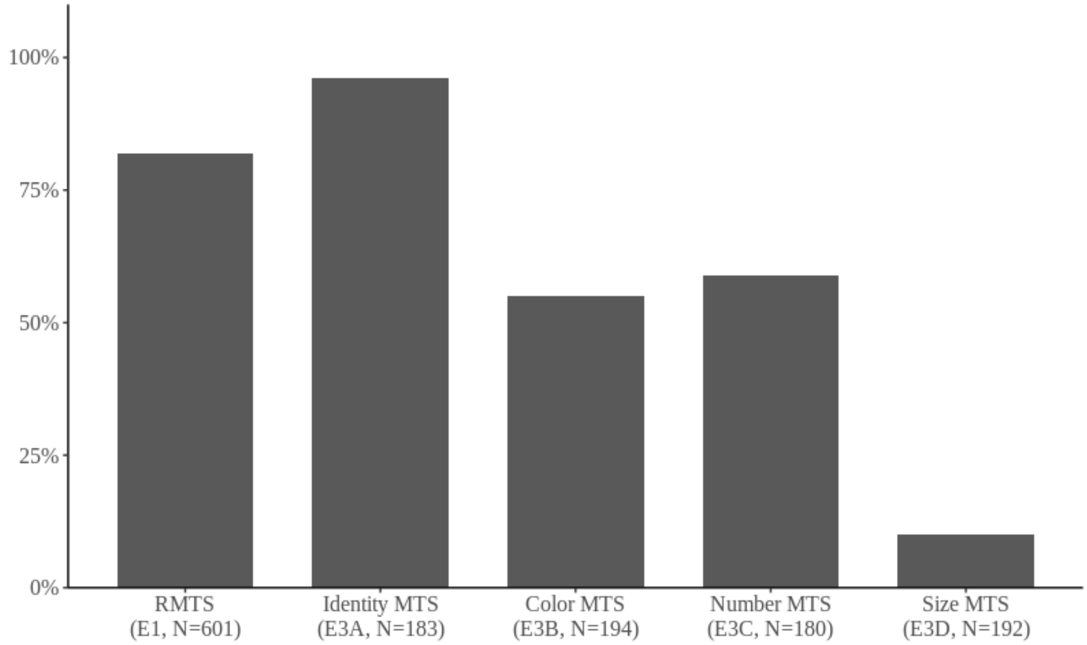


Figure 4. Percentage of adults spontaneously succeeding (8/8 trials correct, no feedback) on (R)MTS tasks in Experiments 1 and 3A-D (tasks and sample sizes are indicated on the X axis labels).

<u>OMTSvRMTS</u>

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 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, Size). There was a main effect of training condition (No training, Identity MTS, Color MTS, Number MTS and Size MTS; (F(4, 898) = 14.80, p < .0001). Post-hoc tests using Tukey's 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.

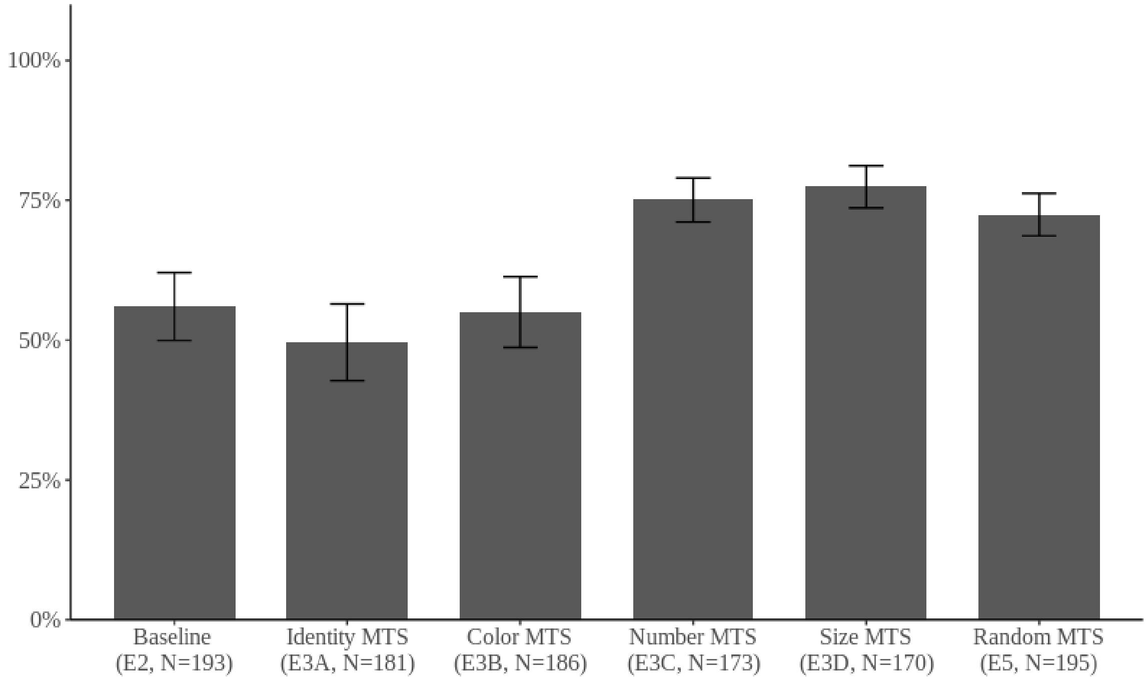


Figure 5. Percentage relational matches on OMTSvRMTS with no training (Baseline, Experiment 2), and after training on Identity MTS, Color MTS, Number MTS, Size MTS and Random MTS (Experiments 3 and 5). Sample size excludes participants who did not succeed above chance on MTS training. Error bars display standard errors.

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[8]. 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; Vendetti, Wu & Holyoak, 2014; Jamrozik & Gentner, 2020). However, to our knowledge, this is the first demonstration of a facilitation of relational matching via a training experience (MTS tasks) which 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, Wasserman & Young, 2001 and according to a reviewer for Kroupin & Carey, under review, *most* non-human animals in RMTS studies have had previous experience with Identity MTS). On the other hand, there is reason to suppose

[8] 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.

the MTS 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 non-differentially 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, prior to 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. Vendetti, Wu & Holyoak, 2014; Simms & Richland, 2019). These studies gave some participants (adults, Vendetti et al., 2014; children, Simms & Richland, 2019) experience with completing analogies - a second-order relational reasoning task - prior to 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 pre-existing inductive biases of the population relevant to the task at hand - here US adults and OMTSvRMTS, respectively. Second, the mechanisms by which these biases are changed by training - here MTS tasks in Experiment 3.

Pre-existing inductive biases: The pattern of spontaneous success on RMTS (Experiment 1) and MTS tasks (Experiment 3; see Figure 4) show adults’ pre-existing 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[9]). 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 (which 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' pre-existing 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, i.e. 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' Pre-existing 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.

[9] 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.

Identity MTS

The near-ceiling rate of spontaneous success on Identity MTS suggests that this task is almost perfectly in line with the pre-existing 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.

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 therefore 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 (which 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 since 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 (since all cards have two objects), conceiving the criterion in the second way would increase the likelihood of inferring relational matches as correct on OMTSvRMTS, (since 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 (which 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' pre-existing 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 pre-existing 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

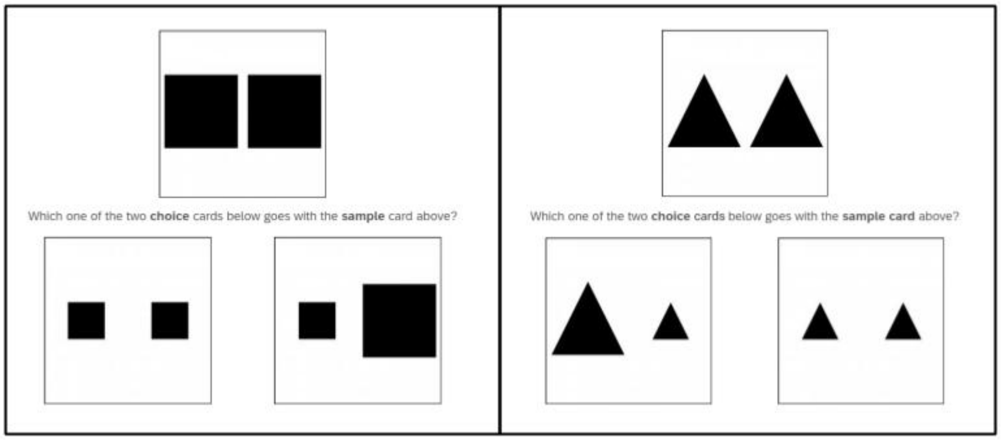
Our hypothesis is that adults who do not match on relations in OMTSvRMTS have inductive biases which lead them to prefer matches on *shape and/or color specifically*. It follows that if we make shape and color less salient as matches in OMTSvRMTS, adults should become more likely to match on the relation 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 v. 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 which adults are extremely unlikely to spontaneously infer as the correct basis of matching (10% spontaneous success on Size MTS, Experiment 3; Figure 4).

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 (Vendetti et al., 2014; Simms & Richland, 2019), there should be no difference between performance on this modified OMTSvRMTS task and baseline performance (Experiment 2). In sum, 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* towards 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.



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?

Figure 6. Two trials of the modified OMTSvRMTS task.

Participants

We recruited 204 participants from Amazon Mechanical Turk who had not participated in any RMTS study from our lab (mean 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 matches or relational matches *in general*, but rather 2) over matches on shape and/or color v. 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 which assumes that preference for relational matches is established at the level of all relations v. all object properties (e.g. Vendetti et al., 2014; Simms & Richland, 2019) can explain how the proportion of relational matches can go from 53% to 93% without the relational structure of the task changing *whatsoever*.

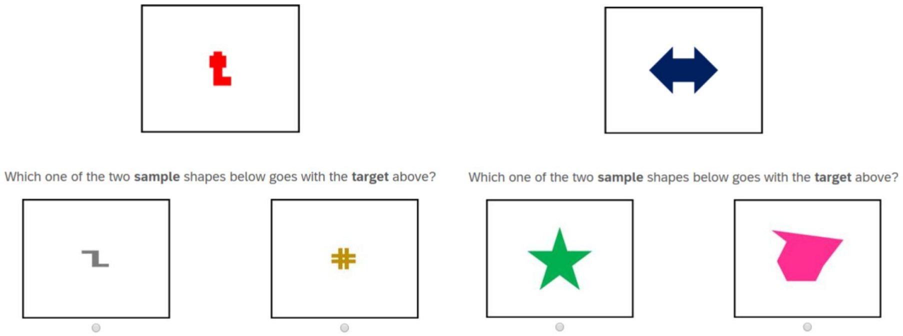
Experiment 5

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 which 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* is irrelevant; it just can't 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 which has partial shape/color matches available (i.e. stimuli vary on shape and color) but has *no correct basis of matching at all*, i.e. error feedback is randomly assigned for each trial. We call this a Random MTS task (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.



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?

Figure 7. Two trials of Random MTS.

Participants

We recruited 196 participants via MTurk who had not participated in any of our MTurk studies on RMTS (mean 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 *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 to 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, pre-existing 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, i.e. 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 pre-existing 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 v. non-human animals, young children) may be due to inductive biases alone.

1) 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., under review). 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.

2) 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, under review) and 4 (Kroupin, 2020) with four-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, Dobson, Esper & Pearce, 2010).

3) 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 which 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, i.e. leading participants to attempt and then inhibit shape/color matches.

Possible effects on inductive biases in previous paradigms

Looking back at previous training paradigms which 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; Thomspson, Oden & Boysen, 1997) may have made pre-existing 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 pre-existing 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 which 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 to 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 pre-existing 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 non-human animals by changing inductive biases *alone* remains an empirical question - one which 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 non-human animals were initially trained on Identity MTS (e.g. Cook & Wasserman, 2007; Fagot, Wasserman & Young, 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 thus making RMTS success *more difficult*. Evidence from categorization-learning studies with non-human 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* non-human animals have experience with Identity MTS prior to 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 - i.e. 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. (2021) 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 which 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 towards 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% v. 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 & Carey, under review; Kroupin, 2020).

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 eight and twelve 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 pre-existing 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, Gay, Glick & Sharpe, 1971, p. 22)

APPENDIX TO PAPER 1

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: Mean = 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 incolor 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: Mean = 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: Mean = 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: Mean = 35.72, SD = 11.21

CHAPTER 3

You Cannot Find what You are not Looking for: Population Differences in Relational Reasoning are Sometimes Differences in Inductive Biases Alone

Introduction

Relational reasoning, including the ability to align relations across different sets of individuals, underpins many of our proudest achievements as a species. Art depends on metaphor, science depends on analogies, mathematics is *nothing but* relations and our everyday language is saturated with representations of relations (e.g. Holyoak & Thagard, 1995; Halford, Wilson & Phillips, 2010; Kotovsky & Gentner, 1996). Without relational reasoning we would neither be able to make, nor comprehend, analogies like those between atoms and solar systems - as Bohr (1913) did in his famous model of the atom, or metaphors that compare the emergence of a teenager onto a balcony with the rising of an vast ball of burning hydrogen over the horizon - as Shakespeare (1595) did in *Romeo and Juliet*. A fundamental question for cognitive science, therefore, is: How do humans come, over phylogeny and ontogeny, to perform these feats of relational processing?

Population differences in relational reasoning - evidence from RMTS

Premack (1983) introduced Relational Match to Sample (RMTS) as a test of whether a given population was capable of relational reasoning *at all*. RMTS has since become the 'gold standard' in assessing basic relational reasoning abilities (Christie & Gentner, 2014). In RMTS the participant is presented with three pairs of objects - in most cases geometric figures displayed on cards (see Figure 1). The figures within a pair can either be identical (same-figure cards) or distinct (different-figure cards). One card serves as the sample and the two others serve as choices. The correct choice card is the one which instantiates the same relation as the sample - same goes with same, different goes with different. This task requires relational reasoning because it involves a mapping of *relations* across two sets, where the individuals in the aligned sets differ from each other.

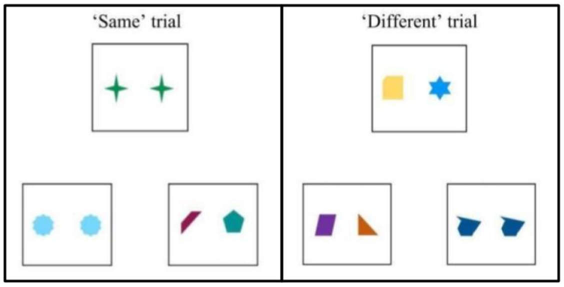


Figure 1: Examples of two RMTS trials.

Even in the case of such an apparently simple instance of relational reasoning (matching same to same and different to different) a striking gap in performance is evident between humans above the age of five or so and all other populations. That is, while older children and adults spontaneously succeed on RMTS (see Hochmann et al., 2017 and Kroupin & Carey, 2021 for child and adult data, respectively), non-human animals and younger children generally fail - despite correct/incorrect feedback for eight trials in the case for children and up to 60,000 trials for non-human animals (see Hochmann et al., 2017 for evidence of failures of four-year-old children despite training involving error feedback and Wasserman et al., 2017 for a review of the comparative literature).

Plan of the paper

The current paper helps discriminate between competing accounts of the sources of such striking population differences in RMTS performance (and relational reasoning more generally) in the case of four-year-old children, a population which ordinarily fails RMTS (e.g. Hochmann et al., 2017, Experiment 1, below), contrasted with adults, who spontaneously and easily succeed (Kroupin & Carey, 2021). Specifically, we test whether the population difference between four-year-olds and adults (in US samples) on RMTS performance is, at least in part, a difference in *inductive biases alone*. That is, we test the possibility that this population difference is *neither* (wholly) due to differences in the *capacity* to engage in relational reasoning, *nor* even in the *availability* of same/different representations. Rather, we propose, four-year-olds, unlike adults (in our population), may simply not infer sameness/difference as the correct bases of matching in RMTS *despite being able to succeed on the task*. We review evidence relevant to the various accounts of population differences in the next section, after which we detail previous work which motivates the methods and hypotheses of the following experiments.

Assessing accounts of population differences in relational reasoning

The literature on relational reasoning has offered two classes of explanations for failures of a given population on a given relational reasoning task, such as RMTS: First, populations may differ in the *capacities* required to be able to engage in relational reasoning at all, and, second, they may differ in the *learning experiences* required to actually generate and use the particular relational representations necessary for the success on the task in question (i.e. sameness/difference in RMTS).

Capacity limitation accounts of population differences in RMTS performance

Limited representational capacity: Numerous authors (e.g., Penn, Holyoak and Povinelli, 2008) have proposed that non-human animals and young children lack the *representational capacity* to form relational representations altogether, in this case of the relations same and different, "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).

Limited computational capacity: Another capacity-limitation account proposes that population differences in relational reasoning are differences in *computational capacities*, such as the presence of sufficient working memory slots to hold in mind the relational comparands (e.g. Halford, 1993). In practice, the computational demands of RMTS in fact vary as a function of the nature of same/different representations used to complete the task (e.g. Thompson et al., 1997): Any mechanism for relational comparison which requires holding in mind representations all of the individual objects in the sets being compared (e.g. the four objects A, A and B, B) poses a much greater working memory challenge than if the relations are encoded as single symbols (e.g. the words “same” and “different”, requiring only one working memory slot each).

Evidence bearing on capacity limitation accounts

Multiple lines of evidence have demonstrated that capacity limitations *cannot* be responsible for all population differences in RMTS performance: Non-human animals have been led to succeed on standard RMTS in a number of training paradigms. These include ‘dogged training’ (Premack, 1983) of tens of thousands of reinforced trials (e.g. Fagot & Thompson, 2011) and training to map the same/different relations on to external symbols (e.g. Premack, 1983; Thompson, Oden & Boysen, 1997). Wasserman, Castro and Fagot (2017) review both dogged and symbol training literatures.

Training studies have also shown that children under the age of spontaneous success on standard RMTS (i.e. five or later, Hochmann et al., 2017; Premack, 1983) can succeed on a variety of modified RMTS tasks, albeit ones which do not involve matching on same *and* different relations (e.g. Kotovsky & Gentner, 1996; Christie & Gentner, 2014) and can certainly successfully discriminate between same and different pairs (Walker & Gopnik, 2014) even from infancy (Hofmann et al., 2016). Such evidence (discussed further below) suggests that children under the age of five have the capacities to succeed on standard RMTS. However, to date, there have been no published data confirming this possibility. In sum, capacity limitations accounts of population differences in standard RMTS performance have been *ruled out* in non-human animals, and there is strong, albeit not decisive, evidence suggesting this may be the case for children under the age of five as well. Consequently, we require an alternative account of population differences in these cases.

Learning experience accounts of population differences in RMTS performance

Absence of specific representations: One alternative account of failures at RMTS (e.g. Gentner, 1988; Christie & Gentner, 2014) proposes that young children and at least some non-human animals have not had *the learning experiences required to generate representations of particular relations* - in the case of RMTS representations of sameness and difference. That is, while young children and non-human animals may be capable of *generating* the kind of abstract representations of sameness and difference necessary to succeed on RMTS (contra capacity limitation accounts of their failures), they may not have had an occasion to do so, just as a nine-year-old may be capable of generating a representation of *division* but may not have actually done so if the concept has not yet been covered in her math class. For example, children learn the words "same" and "different" between ages three and four (Hochmann et al., under review). These words are arbitrary, non-iconic, abstract symbols for the relations. It is possible that prior to learning these words, children do not have access to representations of sameness/difference in the necessary format to succeed on RMTS - or even that the words initially map to representations which are not in such a format (see below for a discussion of possible formats).

Differences in inductive biases alone: Several recent papers (Carstensen et al., 2019; Kroupin & Carey, 2021, see also Walker & Gopnik, 2014) have explored a final account of differences in relational reasoning across populations. Namely, these authors propose that population differences in relational reasoning performance may result *neither* from capacity difference *nor* even from differences in available representations. Rather, populations may differ as a result of a difference in *learning experiences affecting inductive biases alone*. That is, some populations may not have had experiences that would lead them to infer the relations same and different as the correct bases of responding - *despite being perfectly capable of success on RMTS*. For instance, young children and/or non-human animals may infer that partial shape matches (e.g. 'these two pairs go together because both contain a pointy shape') are correct in RMTS and thus fail the task *despite* having abstract same/different representations available and the necessary executive capacities to actually match on the basis of a shared relation of sameness or difference.

Discriminating between learning experience accounts

The ambiguity of relational training and labelling

The success of non-human animals and young children on RMTS as a result of the above training paradigm rule out capacity accounts of population differences (e.g. Penn et al., 2008) - with a significant caveat in the case of children under the age of five, who have not yet been shown to succeed on standard RMTS. All successful training studies to date, however, have either 1) directly provided experience with relational matching prior to RMTS test (e.g. Kotovsky & Gentner, 1996) or 2) provided labels for same/different relations (e.g. plastic shapes in Premack, 1983; Thompson, Oden & Boysen, 1997, the words "same"/"different" or a novel noun in Christie & Gentner, 2014). Both strategies can be explained by *either* learning experience account: Experience with relational matching can in principle *produce* new relational representations, as can applying new symbolic labels (tokens, words) for relations (see, e.g. Gentner & Hoyos, 2017). An alternative mechanism, however, for *both* experience with relational matching *and* applying labels to these relations is that these procedures may change inductive biases so as to make matching according to relations *more likely to be inferred as relevant* in the context of the task (see Kroupin & Carey, 2021a,b for extended discussions).

The Smirnova et al. (2015) paradigm

The most promising paradigm for distinguishing between these learning experience accounts of population differences was introduced in studies with crows and parrots by Smirnova et al. (2015) and Obozova et al. (2015), respectively. Specifically, training in this paradigm allowed birds to succeed, spontaneously (i.e., with no error feedback), on *three* separate RMTS tasks (same-size, same-color, same-shape). The paradigm deployed two distinct forms of training. One was what is known as a 'progressive alignment' procedure (Kotovsky & Gentner, 1996) which was built into the RMTS testing blocks: Prior to every non-differentially reinforced standard RMTS test trial (e.g. GG goes with HI or JJ, Figure 1), birds received three differentially-reinforced trials in which relational matches were *also* object matches (e.g. DD goes with DD or EF). In and of itself, effects of progressive alignment can be explained by either learning experience account, i.e. progressive alignment may serve to produce "new relational abstractions" (Gentner & Hoyos, 2017), but it also may simply direct participants' attention to relational matching as a result of relational matches' constant co-occurrence with reinforced object matches. Regardless, it is not clear whether progressive alignment in fact played a crucial role in birds' success: Birds succeeded *at equal rates* on the progressive alignment and full RMTS trials from the very first testing session (progressive alignment/RMTS performance was 76/83% and 75/72% for crows and parrots, respectively). This suggests that both crows and parrots had been ready to succeed spontaneously on RMTS by the time they had completed the training *preceding* the progressive-alignment/RMTS trials.

The second form of training, which preceded progressive-alignment/RMTS trials, was composed of a series of MTS tasks in each which birds were taught to match on the basis of a particular *non-relational* features - first color, then identity[10], then number and finally size (see Smirnova et al., 2021 for a detailed description of this procedure). Given the evidence above that progressive alignment may have had a limited role in birds' RMTS success, it is plausible that MTS training was a crucial part of the success of this paradigm. This possibility is uniquely important for disambiguating between the two learning experience accounts (creating a new representation of a previously unencoded relation vs. changing inductive biases alone). This is because, since unlike *all* previous training paradigms with non-human animals and young children, MTS training neither 1) directly provided experience with relational matching prior to RMTS test nor 2) provided labels for same/different relations. As a result, it is difficult to imagine how such training could produce *new* relational representations. Consequently, evidence that MTS training *by itself* increased relational responding would provide evidence that the effects of such training were to change inductive biases *alone*.

Evidence that MTS training changes inductive biases in adults

Previous work from our lab (Kroupin & Carey, 2021) strongly supports this possibility: Training on some *but not all* MTS tasks (Number and Size MTS, but not Identity or Color MTS), with the same bases of matching as used in the Smirnova et al. (2015) paradigm increase the likelihood that *human adults* make relational matches in a modified RMTS task where relational matches were pitted against incomplete object matches (e.g. AA goes with BB or AC). Given that human adults *clearly already had* the necessary representations and computational capacities to succeed on RMTS (another experiment in the same paper showed they do so spontaneously),

[10] i.e. matches on all dimensions. In fact the original paradigm described them as 'shape' matches, but the objects on sample and correct choice cards in fact corresponded in shape, color and size.

the *only* mechanism by which MTS training could have increased relational responding in this case was by changing adults' inductive biases such that they became more likely to infer relational matches as correct.

The present studies

The goal of the present studies is to establish whether MTS training tasks of the type used by Smirnova et al. (2015) and identical to the ones we used with adults (Kroupin & Carey, 2021) will produce spontaneous RMTS success in a population which ordinarily fails the task - in this case four-year-old children - *without progressive alignment or any other training that directly involves the relations same and different between two individuals*. Our first hypothesis in this study is that the persistent failure of (a significant proportion of) four-year-old children on *standard* RMTS reflects *neither* an absolute capacity limitation (contrary to Accounts 1 and 2) *nor* a lack of the necessary representations (contra Account 3), but a difference in inductive biases *alone* (Account 4). A number of previous results with modified RMTS tasks suggest - though fall short of proving - that four-year-olds (in Western, educated populations) do, in fact, *already have* the abstract same/different representations and computational capacities necessary for success on standard RMTS.

Causal RMTS/Same-different discrimination

In a 'casual RMTS' paradigm, Walker and Gopnik (2014, also see Walker, Bridgers & Gopnik, 2016) have demonstrated that children as young as three succeed in discriminating the relation same from the relation different after only two demonstration trials. In this task, children are presented with a 'blicket detector' - a box which makes noise if the correct items are placed on top of it. Children are then shown that when two identical objects (i.e. a *same* pair) are placed on top of the box it lights up, but not when two distinct objects are placed on top (i.e. a *different* pair). While an important source of evidence, these tasks fall short of evidence of full relational matching since they do not require participants to actually *align* two instances of same or different relations on any given trial, merely to be able to learn a rule ‘choose same’ (or ‘choose different’). In fact, success on other same-different discrimination tasks has been shown with even younger children (e.g. infants, Ferry, Hespos & Gentner, 2015; Hochmann et al. 2016) and non-human animals (e.g. pigeons, see Wasserman et al., 2017 for a review).

Same-different discrimination paradigms provide evidence that certain populations which do not succeed spontaneously on RMTS have *some* abstract representations of sameness and difference. It is not clear, however, whether such representations are *sufficient* for success on RMTS - and even if they are they likely impose a much greater working memory load in doing so. For instance, Hochmann et al. (2016) propose that infants may discriminate same-pairs from different-pairs by mapping any given pair onto a representation of some two particular items held in long term memory, i.e. a representation in the format [X,X]. That is, a child may have formed a long-term memory of two identical items, e.g. a favorite pair of matching cups stored as [cup, cup]. Subsequently, other pairs of identical objects may be aligned with this representation of a *specific* set instantiating the relation *same*. By doing so, the infant can successfully discriminate same-pairs from different-pairs on the basis of ‘aligns with [cup, cup]’ v. ‘does not align with [cup,cup]’. Such representations of sameness and difference plausibly underlie infant and animal success in the Marcus et al. (1999) “rule abstraction” experiments (identifying the similarity among “la di la”, zu mo zu, te pa te” and distinguishing such triads from “di di ga” while generalizing the pattern to “di gu di” (see Hochmann et al., 2016, for discussion of, and evidence for, this possibility).

While this kind of representation is sufficient for success on same-different discrimination, it is far from obvious that it could be used to succeed on RMTS: Relational *matching* using this comparison would involve at a minimum identifying the sample card as 'aligns with [cup, cup]', then identifying one choice card as 'aligns with [cup, cup′]', then identifying these two as aligning *with each other* on the basis of both aligning with [cup, cup]. Even if this operation is *possible*, it requires that *all of the individual objects* in the choice and sample *as well as* internally represented (i.e. [cup, cup]) pairs must be maintained in working memory in order to make this higher-order comparison - a feat that would strain the working memory even of adults.

Progressive alignment

Kotovsky and Gentner (1996) demonstrated that progressive alignment allowed four-year-old children to succeed on a modified RMTS task where the bases of matching were symmetry and monotonic increase (e.g. progressive alignment trials in the symmetry condition would have the form: - aAa goes with bbB or aAa, while a RMTS trial would have the form: xxX goes with yyY or zZz). While this is certainly evidence that children can match over the dimensions of symmetry and monotonic increase, the effect of progressive alignment is ambiguous: One the one hand, it may have served to *develop* representations of these dimensions *de novo* in children. On the other, it may have changed inductive biases so as to make children more likely to infer them as correct bases of matching.

Labelling relations

Christie and Gentner (2014) explored a final pair of paradigms which has produced success on a modified RMTS task by children under the age of five: The first paradigm involved training children to label cards with same or different pairs of objects with the words "same" and “different”. As a result of this training, children as young as three were able to succeed on an RMTS task in which the sample pair was *always* same. The authors argue that this success comes as a result of children developing representations of sameness and difference *for the first time* as a result of this training. An alternative explanation is that (at least some of) these children *already had* such representations available and the training simply changed inductive biases so as to make these relations highly salient as possible bases of matches in the subsequent simplified RMTS task.

The second paradigm used by Christie and Gentner (2014) was to label the sample card with a novel label (i.e. “this is a Truffet”) and then ask children to select the sample card that the label could also apply to (i.e. “which of these is also a Truffet?”) In this paradigm, children as young as two-and-a-half succeeded on the same simplified RMTS task described above. The authors argue that labelling promontory relational comparison, and thus the abstraction of a new relation. Notice that this implies that children *had no* pre-existing abstract representation of sameness which they could use to succeed on the task and then developed it *de novo* within a few trials as a result of applying the word “Truffet” to both sample and choice cards. An alternative possibility is that children mapped the word “Truffet” to a pre-existing representation of sameness. This is all the more plausible since “Truffet” is a singular noun which is being applied to a *pair* of objects - implying that whatever it means must apply to the pair as a whole. Notice that the representation of sameness mapped to “Truffet” could just as readily be the [X,X] format discussed above as a unitary symbol (such as the word “same”). Once “Truffet” is mapped to a representation of sameness, the task becomes a same-different discrimination task: Children can simply choose which of the two sample cards corresponds to “Truffet” on every trial without matching it to the sample card (the card displaying the same-pair is always correct since the sample card *always* displays a same-pair).

In sum, while a series of previous results with modified RMTS tasks suggests that spontaneous success on standard RMTS is likely to be possible for four year olds simply by changing their inductive biases, none of the paradigms used thus far allow us to definitively draw this conclusion. The present experiments fill this gap in the literature by training four-year-olds on MTS tasks drawn from our work with adults (using the same bases of matching as those in Smirnova et al., 2015). In doing so we also test a second hypothesis, namely that there is some continuity in the mechanisms by which MTS training tasks affect relational responding across four-year-olds and adults. This will be evident if the *pattern* of which training tasks do or do increase relational responding is the same in both populations (i.e. Number, Size MTS do, Identity MTS does not).

To recap: Hypothesis 1) Four-year-olds *already have* the necessary same/different representations and capacities to succeed on RMTS and at least some MTS training tasks will change these biases so as to increase the likelihood that they infer these to be the correct bases of matching in RMTS. Hypothesis 2) The same MTS training tasks will and will not increase relational responding in four-year-olds as did so (or not) in adults in Kroupin and Carey (2021), i.e. Number and Size MTS will, but Identity MTS will not. This will provide evidence that the mechanisms by which MTS training tasks have this effect are at least somewhat continuous across development.

Experiment 1

Experiment 1 establishes four-year-olds' baseline performance on RMTS with the particular RMTS stimuli used in these studies, which were also used in Experiment 1 with adults in Kroupin & Carey (2021). Given the failure of children at this age on RMTS in previous studies (e.g. Hochmann et al., 2017; Premack, 1983) we predict that children will not perform above chance.

Participants

Participants were 24 English-speaking children aged 49 to 60 months (M = 54.23m, nine girls, fifteen boys) recruited by phone from the greater Boston/Cambridge area or at a local science museum. One additional child participated in the study but was excluded due to not having sufficient command of English to understand the instructions. The children were drawn from a predominantly middle-class population. All children received a small prize and a high-five for participating. Families who were tested in the lab were also given five dollars of travel compensation.

Materials

Each RMTS trial contained three laminated paper cards, each of which contained two geometric figures (Figure 1). Unique figures were used on every trial. On each trial two choice cards were placed level with each other and below a sample card. The figures on one of the choice cards were the same, and on the other they were different, with the left-right arrangement of the same-figure cards and the different-figure cards counterbalanced across the eight trials. On four trials the sample card was a same-figure card and on four trials it was a different-figure card. The correct choice card was the one on which the two figures stood in the same relation as those on the sample card. The composition and arrangement of individual triads were the same across all participants, but the order in which the triads were presented was randomized - with the constraints that the task did not begin with more than two consecutive same-figure sample cards or more than two consecutive different-figure sample cards and also that there were never more than three consecutive same-figure sample cards, or different-figure sample cards in a row.

Procedure

Children were told that they would be playing a matching game. Choice cards were produced first and placed on the table as the experimenter said “Which one of these two cards…”, then the sample card was produced and placed on the table as the experimenter finished the question “...goes with this card?” After children selected one of the two choice cards, the next trial was presented. No feedback of any kind was given during RMTS trials.

Results

Experiment 1 probes for spontaneous success (no training, not even error feedback) on RMTS. Since there was no feedback of any kind, spontaneous success is at least seven out of eight trials correct, significantly above chance on a binomial test (p = .04). The results of Experiment 1 are consistent with previous work finding failure on standard RMTS tasks by children below the age of five (Premack, 1983; Hochmann et al., 2017). One out of 24 children succeeded spontaneously by this criterion. The proportion of spontaneous succeeders in the sample was not statistically greater than chance on a second order binomial test (p = .35).

In fact, as a group, children made fewer relational matches than would be expected by chance (41% v. 50% relational matches, p = .03, t = 2.24, Figure 4). This below-chance performance is inconsistent with previous studies using the RMTS paradigm with this age in which children perform at chance as a group (e.g. Premack, 1983; Hochmann et al., 2017). We ran a series of exploratory analyses which indicated that children’s performance on a single triad was driving the below-chance result. However, these analyses also indicated that this triad was not *consistently* below chance across identical RMTS test tasks in Experiments 2A-C, suggesting that this below-chance performance was not robust[11]. In any case, the RMTS test task was identical throughout all experiments, and the presence of a triad that aligned with existing inductive biases so as to sometimes yield the choice of the non-relational match works against the hypothesis that MTS training will increase relational matches. Consequently, results remain directly comparable across Experiments 1 and 2A-C regardless of a possible bias within a single triad against relational matches. Nonetheless, because we are testing the hypothesis that training on the MTS tasks changes inductive biases such that the child will successfully match on the relations same and different, we must ensure that differences between training groups is not driven by a statistical difference between below chance performance and merely chance performance. Consequently, performance on RMTS in Experiment 2, which explores the effects of training tasks on a subsequent RMTS task, is compared *both* to baseline performance in Experiment 1 *and* to chance (50% correct) directly.

Experiment 2

In Experiment 2, four-year-olds were trained on one of three MTS tasks - Number MTS (Experiment 2A), Size MTS (Experiment 2B) or Identity MTS (Experiment 2C) - prior to being tested on RMTS. Given any effects of MTS training tasks on a subsequent RMTS test would require success on the former, and given the possibility that young children would not succeed given only correct/incorrect feedback, we provided full instructions regarding the correct basis of matching when a participant made an incorrect choice on an MTS task (the same procedure as

[11] Participants in Experiment 1 deviated from chance performance on only one triad out of the eight included in the task (17% v. 50%, t = 2.56, p = .01), while performance on all other triads did not differ from chance (t ≥ 1.7, p ≥ .1). An ANOVA examining the effect of stimuli triad on responding found a main effect of triad F(7,184) = 2.11, p = .04), and a Tukey HSD post-test indicated that this was driven by a significant difference between the below-chance triad and another triad (17% v, 63% correct, t = 3.60, p < .001). Performance on the latter triad did not differ from chance (63% v, 50% correct, t = .86, p = .39). In order to establish whether the below-chance triad differed consistently from others we ran ANOVAs on results on RMTS from Experiments 2A-C as well: There was a significant effect of triad in Experiment 2A F(7,184) = 2.46, p = .02), driven by the same single contrast as in Experiment 1. There were no effects of triad in Experiment 2B F(7,184) = 1.24, p = .28) or 2C F(7,184) = 1.65, p = .12). Thus, this particular triad did not consistently lead to lower performance than other triads across experiments.

used with adults in Kroupin & Carey, 2021). This feedback helps guarantee that participants succeed on each MTS task on the basis for which it was designed (e.g. number matches in Number MTS). Furthermore, rates of success without correction on MTS tasks (i.e. the proportion of children choosing 8/8 correct) will provide some indication of what bases of matching children's *pre-existing* inductive biases lead them to infer as correct.

If *any* of the MTS training tasks lead children to succeed on RMTS, this will support Hypothesis 1: Four-year-olds *already have* the necessary representations and computational capacities to succeed on standard RMTS. If, moreover, Number and Size MTS training have this effect, but Identity MTS does *not* (i.e. the same patterns as seen with adults in Kroupin & Carey, 2021) this will support Hypothesis 2: The mechanisms by which MTS training tasks change inductive biases are at least somewhat continuous between four-year-olds and adults.

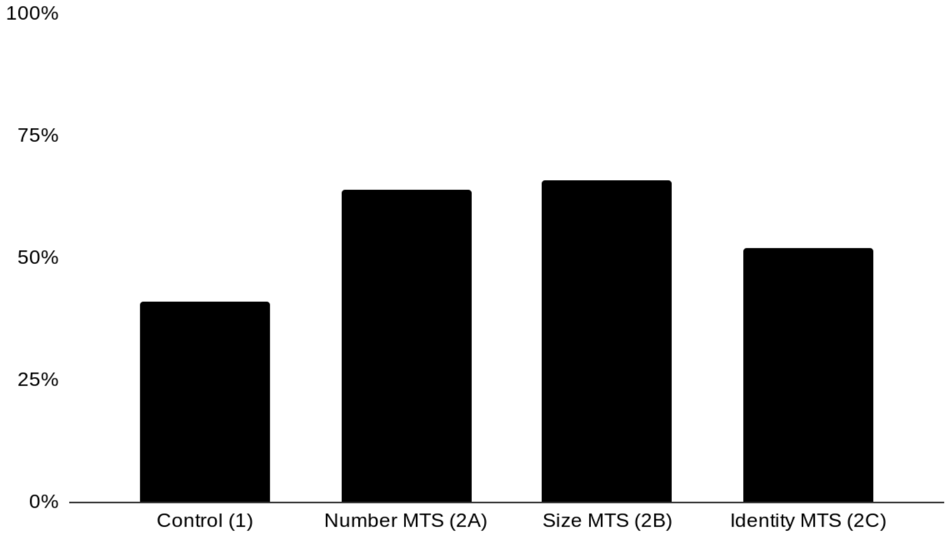


Figure 4. Overall percentage correct on RMTS trials by children in each of Experiments 1-2C .The x-axis displays the training task used (if any) with the experiment number in parentheses.

Experiment 2A: Number MTS training

Participants

Participants were 24 English-speaking children aged 48 to 60 months (M = 53.57m, twelve girls, twelve boys). Demographics and compensation were the same as in Experiment 1. One additional child participated in the study but was excluded due to an insufficient command of English to understand instructions.

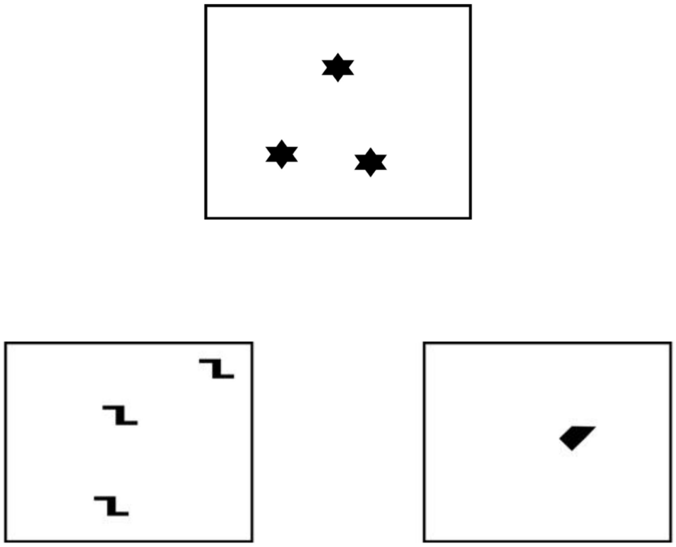


Figure 5: Number MTS trial

Materials

<u>Training task - Number MTS</u>: Each Number MTS card contained one or three geometric figures (Figure 5). All figures were of the same color (black) and there were no figures in common across cards (though figures within each card were identical). On each trial two choice cards were placed level with each other and below a sample card. One choice card contained three figures, the other contained one figure. Four sample cards contained three figures and on

four trials the sample card contained one figure. The correct choice card was the one which contained the same number of figures as the sample card. Sample and choice triads were the same across participants. Their order was randomized, subject to the same constraints on order as in Experiment 1.

Test task - RMTS: The RMTS stimuli were identical to those in Experiment 1.

Procedure

Training task - Number MTS: Children were told that they would be playing two matching games. Children first completed eight trials of Number MTS. Choice cards were produced first and placed on the table as the experimenter said "Which one of these two cards…", then the sample card was produced and placed on the table as the experimenter finished the question "...goes with this card?" If the correct card was selected, the next trial was administered. If the participant chose the incorrect card, the sample card was placed next to the incorrect choice card with the explanation "In this game these two cards *don't* go together because this one has three pictures and this one has one picture" (or vice versa, as appropriate). The sample card was then placed next to the correct choice card with the explanation "In this game *these* two cards go together because this one has three pictures *and* this one has three pictures" (or 'one picture *and* this one has one picture', as appropriate). After the correction was issued the next trial was administered. Once Number MTS was completed the experimenter indicated that the first game was finished and that now the second game would start.

Test Task - RMTS: Eight trials of RMTS were then administered with the same procedure as in Experiment 1, with children receiving no feedback of any kind, not even error correction.

Results

Training task - Number MTS: Overall children chose correctly on 86% of the total Number MTS trials. Needless to say, this was significantly above chance performance (p < .0001, t = 16.23) - as were children's performances on all MTS tasks presented here. Such above-chance performance on MTS training tasks is hardly shocking given that children were given explanations of the correct basis of matching every time they made an error on an MTS training task. Consequently, we do not report further overall performance on MTS training tasks.

Notice that the criterion for spontaneous success is slightly different between RMTS test and MTS training: Since children did not receive corrective feedback in RMTS, performance above chance (i.e. 7+/8 trials correct) was considered spontaneous success. In contrast, since children receive corrective feedback on MTS training tasks if they choose incorrectly, only 8/8 trials correct is spontaneous success.

Four out of 24 (17%) children succeeded spontaneously on Number MTS, i.e. chose correctly on every trial and were never corrected. This proportion is greater than would be expected by chance on a second-order binomial test (p < .0001). Nevertheless the relatively small overall number of succeeders is consistent with previous work showing that children at this age are unlikely to spontaneously attend to the number of figures on a card as a basis of matching (e.g. Chan & Mazzocco, 2017). In contrast, over half (59%) of adults succeed spontaneously on all trials of an identical Number MTS task (Kroupin & Carey, 2021). Thus, the inductive biases relevant to Number MTS clearly change between age four and adulthood, at least in our population, confirming previous work (e.g. Chen & Mazzocco, 2017). Regardless, the majority of children in our sample had no trouble completing the task when given correction - only four out of the twenty children who received a correction chose incorrectly on any subsequent trial. Thus, four-year-olds clearly have the *capacity* to match on the basis of number.

Test task - RMTS: In spite of receiving no corrective feedback, not even error correction, and a mere *eight trials* of Number MTS as training, seven out of 24 children chose correctly on seven or eight out of eight RMTS trials and thus were spontaneous succeeders. This proportion was statistically greater than chance on a second order binomial test ($p < .0001$). Likewise, as a group, children in Experiment 2A made more relational matches than would be expected by chance (64% v. 50% relational matches, $p < .01$, $t = 2.91$). In addition, children made significantly more relational matches on RMTS after completing Number MTS than those in the baseline RMTS study (Experiment 1), 64% v. 41% relational matches, $p < .001$, $t = 3.67$, Figure 4. In contrast to previous training paradigms, a mere eight trials of a MTS task in which the correct response was the choice card with the same number of elements on it as the sample card, involved no training on labelling of relations. Nor did it involve relational matching since number is a property of a set, not a relation among individuals within a set. Brief training on just one of the MTS tasks from the Smirnova et al. (2015) training paradigm is sufficient to increase four-year-olds' relational responding on standard RMTS, leading to overall above chance performance.

The results of Experiment 2A constitute the first evidence that a population which ordinarily fails standard RMTS does so at least in part as a result of differences in inductive biases *alone* (relative to older children and adults, who succeed spontaneously). The training experience that led to success, a mere eight trials of a MTS task in which the correct response was the choice card with the same number of objects on it as the sample card, involved no training on relational matches. Nor does Number MTS training involve nor labelling of relations since number is a property of a set, it is *not* a relational property. Eight trials of an identical Number MTS training task also increased relational responding among adults in a modified

RMTS task (Kroupin & Carey, 2021). In the latter adult study the only possible change as a result of training was in inductive biases alone, since US adults clearly already have the specific representations and computational capacities necessary for success on RMTS, evident from their spontaneous success on the task. The adult finding lends further plausibility to the conclusion that Number MTS training led four-year-olds to succeed on RMTS as a result of changing inductive biases alone.

Experiment 2B: Size MTS training

Participants

Participants were 24 English-speaking children aged 49 to 60 months (M = 53.20m, thirteen girls, eleven boys). Recruitment, demographics and compensation were as in Experiment 1. Five additional children participated in the study but were excluded for failing to complete the study or due to parental interference.

Materials

Training task- Size MTS: Each Size MTS card contained one geometric figure which was either relatively small or relatively large (such that relatively large figures were at least three times the height and width of relatively small figures; Figure 6). On each trial two choice cards were placed level with each other and below a sample card. One choice card contained a relatively small figure and the other contained a relatively large figure. On four trials the sample card contained a relatively large figure and on four trials the sample card contained a relatively small figure. The correct choice card was the one which contained a figure of approximately the same size as that on the sample card. Sample and choice triads were the same across participants. Their order was randomized, subject to the same constraints as in Experiments 1 and 2A.

Test task - RMTS: The RMTS stimuli were identical to those of Experiment 1.

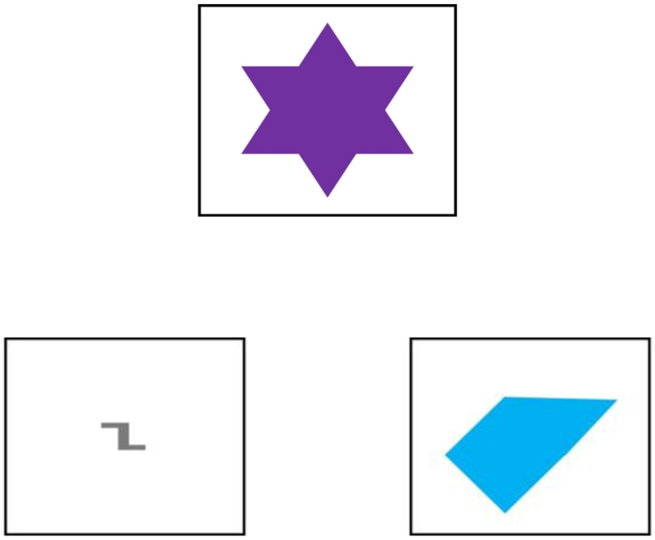


Figure 6: Size MTS trial

Procedure

The procedure for Experiment 2B was identical to that of Experiment 2A, except children completed eight trials of Size MTS instead of Number MTS. The correction procedure in Size MTS in response to an incorrect choice was adjusted accordingly: If the participant chose the incorrect card, the sample card was placed next to the incorrect choice card with the instruction “In this game these two cards *don’t* go together because this one is little and this one is big.” (or vice versa, as appropriate) The sample card was then placed next to the correct choice card with the instruction “In this game *these* two cards go together because this one is little *and* this one is little.” (or ‘big *and* this one is big’, as appropriate).

After Size MTS was completed, RMTS was administered with the same procedure as in Experiment 1, with children receiving no feedback of any kind.

Results

<u>Training task - Size MTS</u>: Only four out of 24 (17%) children chose correctly on every trial of Size MTS and were never corrected. While this proportion is above chance on a second-order binomial test ($p < .0001$), it is nevertheless evidence that children are quite unlikely

to spontaneously infer size as the correct basis of matching. Unlike the case of Number MTS, where children succeeded spontaneously at a lower rate adults, the low rate of spontaneous success on Size MTS is comparable to adults' performance on identical task (17% spontaneous succeeders among four-year-olds on Size MTS in Experiment 2B, 10% spontaneous succeeders on Size MTS for adults in Kroupin & Carey, 2021). Thus, while in contrast to developmental changes in the initial inductive biases relevant to Number MTS and RMTS between age four and adulthood, biases relevant to Size MTS are stable across this age gap - neither population spontaneously infers that size is a sensible basis of matching geometric figures. Failure to spontaneously infer size as the correct basis of matching is perhaps unsurprising since geometric figures have no *canonical* size, unlike real-world objects whose size is encoded automatically in the visual system, even by age four (Long, Moher, Carey & Konkle, 2019). Therefore, there is little reason why individuals should develop a bias to attend to the relative sizes of geometric figures as one of their relevant properties when comparing them to other geometric figures.

Nonetheless, children in Experiment 2B overwhelmingly succeeded on all trials after receiving instructions with only one out of twenty children who received correction choosing incorrectly on any subsequent trial. In spite of inductive biases which lead both adults and children to infer properties other than size as the correct basis of matching in MTS tasks, both groups easily succeed when told that size is the correct basis of matching.

<u>Test task - RMTS</u>: In spite of receiving no correction on the RMTS test trials whatsoever, not even error correction, five out of 24 children succeeded spontaneously, choosing correctly on seven or eight of eight test trials. This proportion was statistically greater than chance on a second order binomial test ($p < .0001$). Likewise, as a group, children in Experiment 2B made more relational matches than would be expected by chance (66% v. 50% relational matches, p <

.001, t = 3.60, Figure 4). Furthermore, as a group, children made more relational matches than those in Experiment 1 (66% v. 41% relational matches, p < .0001, t = 4.19).

The results from Experiment 2B converge with those of Experiment 2A, first in showing unequivocally that training on a MTS task where the match is on an object property can increase relational responding in a subsequent RMTS task in a population that would otherwise perform at chance. This evidence also converges, therefore, in providing the first evidence that a population which ordinarily fails standard RMTS does so at least in part as a result of differences in inductive biases *alone* (relative to older children and adults, who succeed spontaneously). The training experience that led to success, a mere eight trials of a MTS task in which the correct response was the choice card with the same size object on it as the sample card, involved no training on relational matches, nor labelling of relations. Moreover, size is a property of a single object, not a relation between two individuals in a sample array. Eight trials of Size MTS training also increased relational responding among adults in Kroupin & Carey (2021), in a study where the only possible basis of change was one of inductive biases alone (since adults *already* have the necessary representations and computational capacities). The parallel result in our work with adults lends plausibility to the conclusion that Size MTS training led four-year-olds to succeed on RMTS as a result of changing inductive biases alone.

Experiment 2C: Identity MTS training - RMTS test

Unlike Number and Size MTS, Identity MTS did not increase relational responding in adults (Kroupin & Carey, 2021). As a result, the question of whether Identity MTS training significantly increases relational responding in four-year-olds is an important test of whether the mechanisms by which Number and Size increased relational responding in Experiments 2A and B is similar to that by which they did so in adults. If these mechanisms are similar across age, we would expect the pattern of which MTS tasks do or do not increase relational responding to also be consistent across four-year-olds and adults (i.e. Number and Size MTS do, Identity MTS does not).

Participants

Participants were 24 English-speaking children aged 49 to 59 months (M = 53.4 months, ten girls, fourteen boys). Recruitment, compensation, and demographics were the same as in Experiment 1. Four additional children participated in the study but were excluded for failing to complete the study, parental interference or experimenter error.

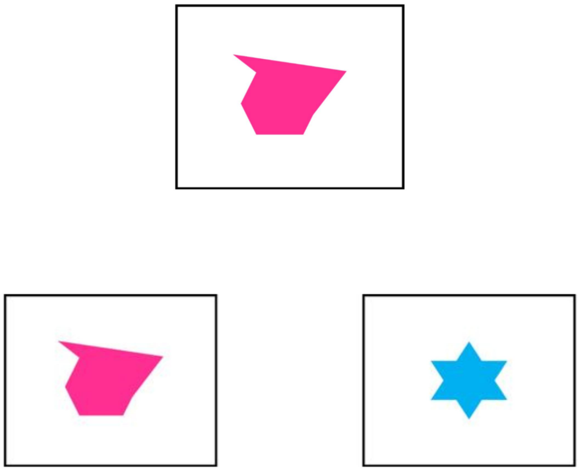


Figure 7: Identity MTS trial

Materials

Training task - Identity MTS: Each Identity MTS card displayed one geometric figure (Figure 7). On each trial two choice cards were placed level with each other and below a sample card. The figures on the two choice cards were different from one another, while the sample card contained a figure identical on all dimensions to one of the two choice cards. The correct choice card was the one which contained the same figure as the sample card. Sample and choice triads were the same in composition and arrangement across participants; their order was randomized.

Test task - RMTS: The RMTS cards were identical to those in Experiment 1.

Procedure

The procedure for Experiment 2C was identical to that of Experiments 2A and 2B, except children completed eight trials of Identity MTS instead of Number MTS or Size MTS. The correction procedure in Identity MTS in response to an incorrect choice was adjusted accordingly: If the incorrect choice card was selected, the experimenter issued a correction: The sample card was placed next to the incorrect choice card with the explanation "In this game these two cards *don't* go together because the picture on this one does not look like the picture on this one." The sample card was then placed next to the correct choice card with the explanation "In this game *these* two cards go together because the two pictures look like each other."

After Identity MTS was completed, RMTS was administered with the same procedure as in Experiment 1, with children receiving no feedback of any kind.

Results

Training task - Identity MTS: The vast majority of children (21 out of 24 or 88%) chose correctly on all trials of Identity MTS and received no correction. Needless to say, this proportion of spontaneous success is greater than would be expected by chance on a second-order binomial test ($p < .0001$). The three remaining children made only one error and received only one corrective explanation. Clearly, the inductive biases of children are highly likely to lead them to infer identity matches (i.e. matches on shape, color and size) as the correct basis of matching. Furthermore, the contrast between the overwhelming spontaneous success on Identity MTS and lack thereof in Size MTS suggests that children's inductive biases lead them to *specifically* infer shape and color to be correct bases of matching. This result is consistent with a large literature showing matches on shape and/or color to be highly salient to children of this age (e.g. Gentner & Ratterman, 1991; Richland, Morrison & Holyoak, 2006; Chan & Mazzocco, 2017; Landau, Smith & Jones, 1992). These data are furthermore in line with adults' performance on an identical task; adults also overwhelmingly (96%) succeeded on all eight trials with no correction (Kroupin & Carey, 2021). There is continuity between age four and adulthood in the inductive biases relevant to Identity MTS (high spontaneous success) and Size MTS (low spontaneous success), most probably due to both populations' inference that shape and/or color are the correct bases of matching geometric figures. This continuity from the preschool years to adulthood is *not* seen in the rates of spontaneous success on Number MTS and RMTS (markedly lower rates of spontaneous success on both tasks by preschool children), consistent with a large literature showing increased spontaneous attention to numerosity across age in Western samples (e.g. Chen & Mazzocco, 2017; MacMullen, Vershaffel & Hanula-Sormunen, 2020).

Test task - RMTS: Two out of 24 children chose correctly on seven or eight out of eight trials and were thus considered succeeders. This proportion was not statistically greater than chance on a second order binomial test (p = .23). At a group level, children in Experiment 2C did not make more relational matches than would be expected by chance (54% v. 50% relational matches, p = .3, t = 1.03). Their performance was statically better than the baseline performance in Experiment 1 (54% v. 41% relational matches, p = .03, t = 2.31, Figure 4), but given the chance-level performance in Experiment 2C, this is clearly a result of the below-chance performance of children in baseline (Experiment 1).

Results of Experiment 2C are consistent with the hypothesis that mechanisms by which inductive biases are changed are at least partially continuous across development: Children's relational matching was facilitated *specifically* by those MTS tasks which also increased relational matching in adults (Number and Size MTS) while Identity MTS, which did not increase relational matching in adults, did not improve children's RMTS performance to above chance either.

General Discussion

Previous results with modified RMTS tasks suggested that merely changing four-year-olds' inductive biases in a matching task might lead to spontaneous success on standard RMTS. However, no previous paradigms allowed us to definitively draw this conclusion. The present experiments fill this gap in the literature by generating three clear results. First, at baseline, when tested on RMTS alone, four-year-olds failed at RMTS, that is failed to make relational matches at levels better than chance. Second, **eight** trials of training on either Number MTS or Size MTS led to above chance performance on a subsequent RMTS task. Third, eight trials of training on Identity MTS did *not* lead to subsequent success on RMTS.

Unlike in the Smirnova et al. and Obozova et al. studies which inspired the current work, children in Experiment 2 received no progressive alignment trials as part the RMTS test task. They also received no demonstration trials for the RMTS task itself, nor any labels for the relations same and/or different (as in Christie & Gentner, 2014). Nor did they receive error feedback, though Hochmann et al. (date) showed that 8 trials of differential feedback on standard RMTS does not move 4-year-olds away from random responding. Thus, the present experiment cleanly establishes that MTS training *alone* can lead to spontaneous success (no error feedback) on standard, full, RMTS in a population that otherwise fails (four-year-olds; Hochmann et al., 2017; Premack, 1983; Experiment 1, above). Our interpretation, throughout this paper, has been that this pattern of results supports the conclusion that the population difference in RMTS performance between four-year-olds and older children/adults is one in inductive biases *alone*. However, as discussed above, the are three alternative accounts of population differences in the literature: 1) Differences in representational capacity, 2) Differences in computational (e.g. working memory) capacity, 3) Differences in learning experience such that representations of sameness/difference necessary for RMTS success *have not yet been* generated. Can any of these three provide an alternative explanation for our results?

Possible alternative explanations for the effects of MTS training

The question concerning whether the training experience given in Experiments 2A or 2B might have changed basic underlying capacities for relational reasoning very clearly must receive a negative answer. *Eight trials* of Number/Size MTS training certainly could not change underlying *capacities* for generating relational representations (Penn et al., 2008), for manipulating relational representations in working memory (Halford, 1993). The literature on training regimes for increasing children's executive function, for example, shows such increases to be attainable, but over weeks, or months, not minutes. And by definition, any actual representation the child draws upon was in the capacity of a child to create.

The remaining possibility is that, according to Account 3, children coming into Experiment 2 *did not have* representations of sameness/difference in a format that could support success on RMTS and that eight trials of Number/Size MTS *produced* these representations *de novo*. This possibility is, however, barely more plausible than that the training led to new representational or computational capacities. Many people have suggested to us, contrary to our claims, MTS training *does* involve representations of the relation *same*. After all, the instructions to match according to Number, Size and object Identity clearly imply that matches should be made on the *same* values of numerosity, approximate size and object identity. Perhaps, the argument goes, drawing upon the symbolic representation of the relation *same* implicated in any MTS task plays a role in constructing a representation the relation same that can support RMTS.

There are two responses to this suggestion. First, if this were the mechanism through which MTS training is affecting RMTS training, then Identity MTS should also lead to success on RMTS, but it did not. Second, MTS need not require a symbol for the concept *same*, just as non-Match to Sample (nMTS) need not require a symbol for the concept *different*. At least since Premack (1983), it has been recognized that the content *same* in MTS sample could be carried by a match computation, the same match computation that is implicated in every act of recognition or categorization. For example, the procedure that underlies successful performance in MTS might be "Store representation of sample, *x*; if subsequently encounter x, choose x." For nMTS the procedure would be "Store representation of sample, *x*; if subsequently encounter x, avoid x." Zentall (2018) and Hochmann et al. (2016) provide evidence that it is exactly these procedures that underlie pigeons and 14-month-old infants' MTS and nMTS performance, respectively. Importantly, there is no mental *symbol* for same or for different in these procedures, only a mental symbol for the sample. The content *same* is in this case implicit, carried by a match computation.

The critical point here is that, the *kind* of representations which support RMTS in principle, and which the task was designed to assess the availability of, are *domain-general* representations of sameness (and/or difference) in a format which allows for comparison across instances of these relations[12]. Most likely such representations would be in the format of unitary symbols (e.g. the word "same" or a non-verbal symbol with the same meaning, which we might write *Ω*). If children in Experiments 2A and 2B *already had* these representations, MTS training tasks could not have *produced* them. If these same children *did not* have these MTS training

[12] It is *of course the case* that children will have generated *some* new representations in the process of the task - at least by virtue of representing novel stimuli. New relational representations in this sense (e.g. *same-pair-of-novel-geometric-figures*) are presumably generated in the face of every novel instance of sameness and consequently not of interest *per se*.

tasks, it is a total mystery why an MTS task which *does not require* such representations would produce them and, moreover, why it would be *specifically this* instance of matching (i.e. Number or Size MTS in our lab) and not one of the vast number of instances of matching they are likely to have experienced in their lives up until participating in this study.

Alternative explanations in the case of Smirnova et al./Obozova et al.

While the proposal that MTS training tasks produced an abstract representation of sameness for four-year-olds in Experiment 2 is untenable, it remains an important possibility in the case of the original work with crows and parrots (Sminova et al., 2015 and Obozova et al., 2015, respectively). This work included many different MTS tasks, and trained flexibility in choosing a relevant dimension in any given triad that satisfied the logic of a matching task, that is, flexibility in finding a dimension on which the choice cards differed and the sample card had a value that matched only one of the choice cards. Testing whether this extensive MTS MTS training was, in fact, sufficient to produce RMTS success in birds is important for many reasons. This could be done by removing the progressive alignment trials from the test blocks. If it were found that the MTS training were sufficient, followup research should examine whether the training led to new representations of the relations same or different, merely taught the birds the logic of matching tasks, or changed inductive biases so as to promote the plausibility that the relations same and different are the correct basis of matching.

The scope of generalization of the present results

While the results of Experiment 2 definitively show that *some* population differences are differences in inductive biases alone, it is important to clarify 1) the degree to which this bears on other population differences in relational reasoning (and RMTS performance specifically) and

2) whether the small number of individual children succeeding in Experiment 2 implies that the *majority* of four-year-olds in fact fail due to a lack of necessary capacities or representations.

Differences in inductive biases definitely do not account for all population differences

Needless to say the finding that some four-year-olds fail RMTS due to differences in inductive biases alone in no way generalizes to *all* population differences. It is overwhelmingly likely that that certain populations lack the representational or computational capacities to engage in relational reasoning (e.g. single-celled organisms), and that in some instances groups fail on a task as a result of not having yet developed the particular relational representations necessary (e.g. responding on the basis of a novel verb meaning "to hold behind your back and then put down", Haryu, Imai & Okada, 2011). In our results specifically, the fact that the majority of four-year-olds in our sample did *not* succeed, regardless of training, leaves open the possibility of failure due to a lack of representational/computational capacities. Likewise it is possible that some proportion of our sample were *capable* of success but had not yet generated representations of sameness/difference in the requisite format.

MTS training is an imperfect mechanism for changing inductive biases

While the present results by no means generalize to all population differences, neither are they reliable evidence that only a small proportion of four-year-olds fail RMTS as a result of differences in inductive biases alone. After all, training given here was, by design, extremely minimal and indirect since we wanted to rule out that the mechanism for change involved changing representational or computational capacities. Identical training did not lead all adults to rely on relational matches either (Kroupin & Carey, in press). Further work should investigate possible reasons why certain four-year-olds failure on RMTS despite Number/Size MTS training in more detail. Moreover, future work should investigate the same issues in younger children:

Children in the population we are studying (generally from middle class, college educated households) learn the words “same” and “different” in their fourth year of life (Hochmann et al, under review). It is possible that spontaneous success on RMTS *depends upon* a unitary symbol for sameness and/or difference, which in children emerges only when they learn the words “same” and “different”. Thus, high priority for future research would be to study three-year-olds, exploring whether MTS training would lead to success on RMTS, and if so, this is only true for children who know the words “same” and “different.”

Investigating inductive biases

Descriptive and explanatory issues

The finding that in some cases population differences in RMTS performance are differences inductive biases *alone* motivates investigating mechanisms underlying the changes to these inductive biases as a result of experience. Specifically, we are faced with three issues: First, the descriptive issue of *how* the inductive biases of four-year-olds in our sample differ from those of older children/adults (in the Western, educated samples), such that they lead the former to fail RMTS without training. Second, the explanatory issue of the *mechanisms by which* Number and Size MTS training tasks change these biases so that four-year-olds (and adults) become more likely to make relational matches in a subsequent task. Third, another explanatory issue is *why* different populations come to have different inductive biases - e.g. the difference between four-year-olds who succeed on RMTS in Experiment 2 and adults (who nearly all succeed on RMTS without training) in Kroupin and Carey (2021). This latter explanatory question, while vital, is outside the scope of the evidence presented in the current study. Experiments 1 and 2 can, however, begin to address the first two issues. Furthermore, Kroupin & Carey (2021) detailed many testable hypotheses concerning the former explanatory issues, and provided two illustrative experiments that supported a specific hypothesis as to the mechanism by which Number and Size MTS changed inductive biases.

Descriptive Issue: Pre-existing inductive biases

To understand *what* changes to children's inductive biases were made by Number/Size MTS training, and the *mechanisms by which* these were made, we first need to establish what these biases were prior to training. Establishing the inductive biases of an individual or population involves studying spontaneous inference. In the case of MTS tasks (including RMTS), we can infer what inductive biases four-year-olds in our sample bring to the table by assessing the rates of spontaneous success (8/8 trials correct, receiving no instructions about the basis of matching) on the four matching tasks preceded by no training- Identity, Number and Size MTS and RMTS. Clearly, four-year-olds' inductive biases are highly unlikely to lead them to infer sameness/difference (Experiment 1), number (Experiment 2A) or object size (Experiment 2B) as the correct bases of matching. In contrast, children are extremely likely to infer matches on object identity as the correct basis of matching (i.e. shape, color and size, Experiment 2C). In the latter case, given children's low rates of matching on size we can infer that four-year-olds' inductive biases particularly favor shape and/or color matches - a possibility supported by previous work on matching tasks with children of this age (e.g. Christie & Gentner, 2010; Chan & Mazzocco, 2017).

Given there are no perfect matches on shape and/or color on RMTS (Figure 1), how can children's inductive biases to match on shape and/or color have led to failure? The issue is only puzzling if we assume that children necessarily look for *perfect* matches between sample and choice cards (i.e. an identical value on one or more dimensions between sample and choice). If this is *not* the case (and we have no evidence that it should be), they may be perfectly happy to make *partial* matches on shape/color, i.e. approximate matches on these dimensions (e.g. 'this card and this card both have pointy shapes', a partial shape match). Notably, children's preference for *partial* shape/color matches in standard RMTS may parallel the preference of around half of adults to prefer *perfect*, but incomplete, shape/color matches over relational matches in the modified RMTS task we used with adults (Kroupin & Carey, 2021). Figure 8 displays the modified RMTS task used with adults, with the right-hand choice card displaying the incomplete object match (and the left-hand choice card displaying the relational match). This helps explain the apparent continuity in the pattern, and potentially of mechanisms, of MTS training task effects across four-year-olds and adults - in both cases training tasks may have changed inductive biases so as to increase the salience of relational matches relative to shape/color matches.

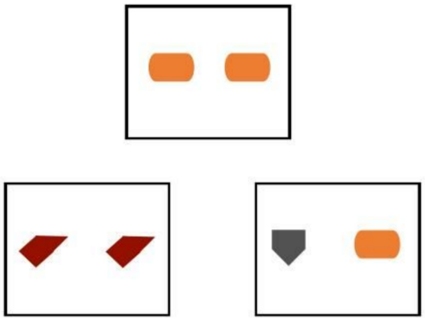


Figure 8. A trial of the modified RMTS task used with adults in Kroupin and Carey (in press)

Explanatory issue: How MTS tasks may affect inductive biases leading to success on RMTS

If children's initial inductive biases lead them to favor partial matches on shape and/or color it is unsurprising that training on Identity MTS would not have affected their choices in a subsequent RMTS task. After all, such training would be almost perfectly in line with their *existing* inductive biases (as we see from near-ceiling spontaneous success) and, as such, unlikely to affect any *change* in them. The crucial question is: How may a mere eight trials of Number or

Size MTS training have affected these biases so as to make them more likely to match on the relations same/different? There are two complementary ways in which any MTS task may affect these biases so as to make matching on relations *relatively* more likely. First, it may inhibit matches on shape and/or color;the MTS training task can increase the likelihood of inferring the relations same and different to be the correct bases of matching by making children *less* likely to infer shape and color as bases of matching, thus leading them to search for other possibilities, including the relations same/different. Second, it may promote matches on the relations same/different - the MTS training task can increase the likelihood of inferring the relations same/different to be the correct bases of matching by making children *more* likely to infer these relations to be the correct bases of matching.

Number MTS

Inhibiting matches on shape and color: For those four-year-olds who succeeded on RMTS after Number MTS training, the latter task may have changed their inductive biases so as to make them less likely to infer shape or color as the correct bases of matching. All figures in Number MTS (see Figure 5) were black, making color an unlikely hypothesis regarding the correct basis of matching in Number MTS and may lead children to infer it is likewise an unlikely basis of matching in the subsequent RMTS task. Furthermore, given that the figures were different across Number MTS cards, children may have initially attempted partial shape matches (e.g.: pictures on this card and pictures on this card are both pointy). This basis of matching would make children highly likely to make an incorrect match at some point during the eight trials of Number MTS and consequently receive a correction indicating not only that number was the correct way to match these cards but also, implicitly, that shape was *incorrect* as a basis of matching. This latter piece of information may make children less likely to search for partial shape matches. By inhibiting initial inductive biases to match on shape and color, Number MTS may thus increase the likelihood that children infer the relations same and different, if these relations are already relatively salient to children (though initially less salient than shape/color).

Promoting matches on the relations same and different: Number is not a property of individual objects, it is a property of sets. Being told that cards with three pictures go with cards of three pictures and cards with one go with cards of one may increase the likelihood that the correct basis of matching should be a property of sets. Since same and different are likewise properties of sets (being relations among individuals with a set) this may increase the likelihood of inferring these relations as the correct bases of matching if these relations are more salient than other possible set properties.

Furthermore, there are two possible ways of interpreting the stated rule in the corrections of errors on Number MTS: Matching by number of figures on the card and matching by number of *the same shaped figures*. After all, figures on each card of Number MTS were identical (consistent with Smirnova et al., 2015). This may lead children to formulate the positive rule ‘match by number of same figures’. Thus, some children who succeeded may have done so by formulating a rule in Number MTS that applies directly to RMTS: Match card with N figures that are identical with a card with N figures that are identical. If this is the case, the relation same may have been facilitated *directly*, changing children’s inductive biases so as to infer the same as a correct basis of matching in a subsequent RMTS task.

These hypotheses are easily tested. For example, If Number MTS promotes matches on the relations same and different by virtue of leading them to infer the rule ‘match by number of same objects’, then making the individual objects within each card on Number MTS have unique shapes should decrease the effect of Number MTS training on subsequent RMTS performance.

Size MTS

Inhibiting matches on shape and color: The most obvious mechanism by which completing Size MTS may have changed four-year-olds' inductive biases so as to make them more likely to infer relations as a basis of matching in RMTS is by providing evidence that shape and color are not correct bases of matching. Given that very few children spontaneously succeeded on Size MTS, it is highly plausible that they initially attempted partial shape or color matches in Size MTS (matching on approximate features of individual objects like both angular, or both light-colored). Such partial shape/color matches would be randomly distributed across the correct and incorrect relational matches, and thus likely to lead to at least one mistake on Size MTS and a correction. The latter not only indicates that size is the correct basis of matching but also, implicitly, that shape and/or color are *incorrect* bases of matching. This, in turn, may make children less likely to search for partial shape and color matches in a subsequent RMTS task, and thus relatively more likely to infer relations to be the relevant basis of matching.

This hypothesis is also easily tested. For example, one could construct a Size RMTS task in which neither shape nor color are available as bases of matching. For instance, a trial could contain a same-figure sample card with two medium-sized black squares, one choice card with one large black triangle and one small black triangle and one choice card with two medium-sized black triangles. All possible color matches are black to black and and all possible shape matches are square to triangle - rendering shape and color uninformative as bases of matching. If four-year-olds can be influenced to succeed on RMTS simply by inhibiting shape and color matches, such a modified RMTS task should allow them to succeed spontaneously (see Kroupin, 2020, for this result).

Effects of MTS tasks: Evidence from continuity

Of course, the above are not the only possible mechanisms through which Number or Size MTS training may have increased relational responding on a subsequent RMTS task. Some mechanisms proposed here do, however, have indirect support: As discussed previously, the pattern of tasks which increased relational responding in four-year-olds was the same as that in adults (Kroupin & Carey, 2021) (Number and Size MTS - yes, Identity MTS - no). This suggests that the mechanisms by which MTS tasks have their effect are at least somewhat continuous across age. Work with adults has provided strong support for the inhibitory mechanisms suggested for Number and Size MTS above: For instance, if inhibiting shape/color matches happens by participants *attempting* these matches and getting feedback that these kind of matches are *incorrect*, this makes the highly counterintuitive prediction that a task with *no right answer* but which allows individuals to attempt partial shape/color matches should increase relational responding. This is, in fact, the case with adults (Kroupin & Carey, 2021). This strongly motivates future research to test whether these inhibitory mechanisms do, in fact, play a role in the effects of MTS training on four-year-olds RMTS performance. It is important to note that this mechanism does *not* concern a domain-general, or even a domain-wide, difference in attention to relations v. object features (as suggested by e.g. Carstensen et al., 2017, Simms & Richland, 2019): This account assumes *some* object features (shape/color) are assumed to be more salient than *some* relations (same/different), while other object features (size) are not (see Kroupin, under review, for further evidence and discussion on this point).

In sum, our goal is to emphasize that some population differences are due to inductive biases *alone*, and that accepting this fact motivates a research program that *specifies* the inductive biases and mechanisms through which these may change.

Looking ahead

We emphasize again that we are not proposing that differences in inductive biases alone can account for *all* population differences in relational reasoning. There will no doubt be cases in which failures result from lack of representational or computational capacities, or the fact that the relevant relational representations have not yet actually been generated.

Instead, these studies highlight two crucial reasons why research on inductive biases should be *included among* the undoubtedly important research on capacity limitations and on the generation of new representations. First, a failure to recognize that populations may differ merely with respect to inductive biases can lead to a misidentification of inductive failures as limitations in the *representations* or *capacities* required to reason relationally. Second, as we can see in the case of RMTS, successful relational reasoning is impossible without successful inference. It follows that characterizing the emergence of the kind of human-unique feats of relational reasoning which motivate this field of study *must also include* a characterization of the emergence of a particular set of inductive biases over phylogeny and ontogeny which support the corresponding inferential processes. In the case of RMTS this means answering the final explanatory question posed by these data and those from previous work, namely: *Why* do young children have inductive biases which lead them to infer bases of matching other than sameness/difference in RMTS, and why does this change by age five or six, in Western populations? We close by repeating a call for, and hope to have shown the feasibility of, integrating *explicit accounts* of the inductive biases, and the development thereof, in the populations from which we draw our samples for relational reasoning research.

CHAPTER 4

Abstract Relations, Particular Biases: Population Differences in Relational Reasoning Sometimes Depend on Inductive Biases Formulated Over Particular Representations

Introduction

Some of the most striking aspects of human cognition - science, literature and even everyday conversation - rely on *relational reasoning*, the ability to compare of abstract relations holding between objects to other such relations holding between other sets of objects (e.g. Holyoak & Thagard, 1995; Halford, Wilson & Phillips, 2010; Kotovsky & Gentner, 1996). Whether or not we are the only population *capable* of relational reasoning, human adults have a clearly unique *facility with* this form of thought: Neither young children nor non-human animals engage in anything like making scientific discoveries through relational comparisons or expressing thoughts in metaphors. It follows that the human facility with relational reasoning depends on changes in *both* phylogeny *and* ontogeny, and that understanding the nature of these changes is a crucial part of understanding human-unique cognition.

RMTS as a measure of population differences in relational reasoning

Relational Match to Sample (RMTS, Premack, 1983) has been referred to as the 'gold standard'(Christie & Gentner, 2014) test of the basic capacities required for relational reasoning since it involves reasoning with arguably the simplest of all relations, sameness and difference (e.g. James, 1890; Wasserman & Young, 2010). A trial of RMTS involves three pairs of objects. Two of these are choice pairs, one of which has identical objects and thus instantiates the relation *same* and the other has two distinct objects, instantiating the relation *different* (Figure 1). The third pair is the sample pair which instantiates either the same or different relation[13]. The correct choice pair is the one whose relation matches the relation instantiated by the sample - same goes with same, different with different.

[13] This is a crucial detail - multiple studies have shown that when *only one* relation is used as the sample the task becomes significantly easier (e.g. Christie & Gentner, 2014; Walker & Gopnik, 2014). These studies are not reviewed here since the simplification involved in these tasks may allow children to succeed without engaging in second-order relational reasoning (see Kroupin & Carey, under review).

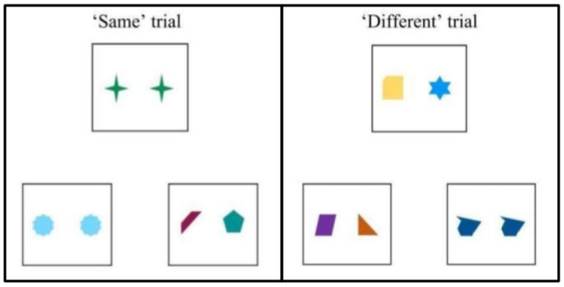


Figure 1: Examples of two RMTS trials.

Performance on RMTS shows significant population differences: Children over the age of five and adults (in Western samples) succeed spontaneously (see Premack, 1983; Hochmann et al., 2017 for evidence from children; Kroupin & Carey, 2021 for evidence from adults). Younger children and non-human animals, in contrast, either fail outright or succeed only with specialized and sometimes very extensive training (see Kroupin & Carey, 2021; Kroupin & Carey, under review; Wasserman et al., 2017 for reviews). In previous work (Kroupin & Carey, under review), we have shown that a mere *eight* trials of training on tasks where objects were matched by numerosity and approximate size (Number and Size MTS), but not on identity (Identity MTS, Figure 2), leads four-year-olds, a population which ordinarily fails standard RMTS, to succeed without any further training on the task. Given the absence of relational matches in MTS training, this result demonstrates that the population difference between four-year-olds and older children/adults is almost certainly one in inductive biases *alone*[14] (see Kroupin & Carey, under review, for the full version of this argument).

[14] Here I mean four-year-olds *as a group*. It is certainly the case that neither the training in our previous work (Kroupin & Carey, under review) nor the studies below allow *all* individual four-year-olds to succeed on RMTS. It remains an empirical whether some four-year-olds lack the capacities or representations necessary to succeed on the task - or whether there is some other intervention/task modification which would allow them to succeed.

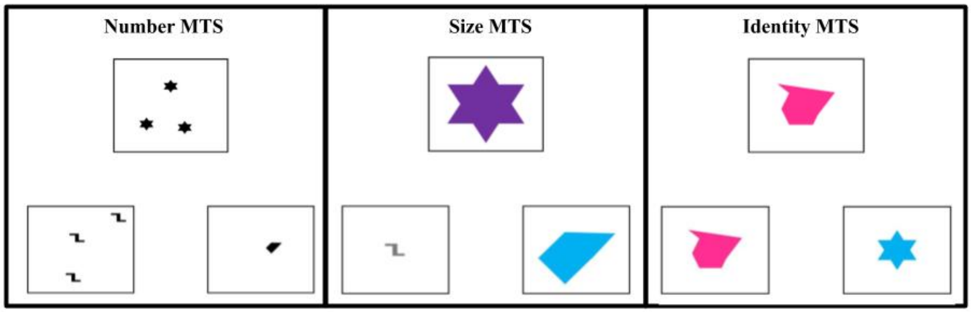


Figure 2: Sample trials from MTS training tasks used in Kroupin & Carey (under review).

This conclusion is further supported by our previous work with adults (Kroupin & Carey, 2021). In those studies, we trained adults on *exactly the same* MTS training tasks prior to testing them on an ambiguous task in which a match on the relation same was pitted against an incomplete object match (Figure 3, left panel). The same MTS training task which increased four-year-olds' performance on RMTS above chance also increased the proportion of relational matches adults made relative to baseline.

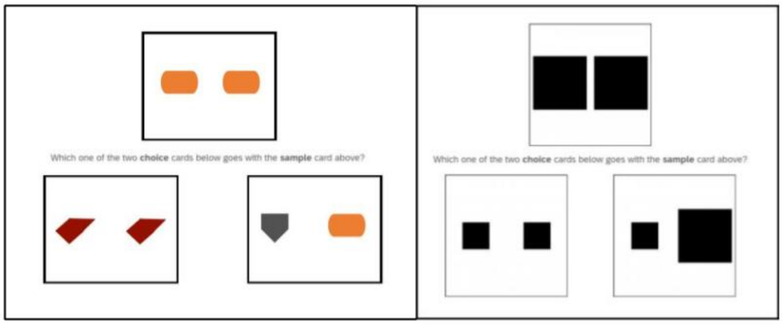

Which one of the two **choice** cards below goes with the **sample** card above?
Which one of the two **choice** cards below goes with the **sample** card above?

Figure 3: Example trials from RMTS tasks used with adults in Kroupin and Carey (under review). Left panel displays the baseline task, right panel displays the task in which shape and color were equated across pairs. In both cases the left-hand choice card displays a match on the relation same, while the right-hand choice card contains an incomplete object match (i.e. one of the two objects matches the objects on the sample card on shape, color and size).

In both papers, we proposed that at least some participants, in both age groups, did not make relational matches as a result of pre-existing inductive biases which led them to infer matches on shape and/or color[15] to be correct. That is, four-year-olds are known to be likely to

[15] Neither our child nor adult studies provide the data to distinguish whether children's and adults' inductive biases lead them to prefer shape alone, color alone or a combination of the two. Consequently I shall refer to the combination of the two as 'shape/color.'

infer shape/color as correct bases of matching in tasks featuring geometric shapes (e.g. Kroupin & Carey, under review, Chan & Mazzocco, 2017). As long as children were content to make *partial* matches on these dimensions (e.g. 'this one and this one go together because they are both pointy'), their pre-existing inductive biases may have led them to prefer such matches over relational matches *even if they were capable of the latter*. The same logic holds for adults: Those adults who did *not* make relational matches in the modified task (Figure 3, left panel) were clearly making incomplete matches on shape/color. That is, their inductive biases led them to infer such matches as more likely to be correct than matches on the relation same.

Again in both papers, we predicted that *reducing* the likelihood of matching by shape and/or color for these populations would increase the likelihood of relational matches. We tested this hypothesis directly in the adult paper by modifying the test task such that shape and color were equated across cards with shapes varying only in size, making the former two dimensions unlikely bases of matching (Figure 3, right panel). Despite the fact that the choices in the two tasks were *identical* (matching on the relation same v. matching one object on shape, color and size), reducing the likelihood that shape/color *specifically* were relevant bases of matching resulted in an increase of total relational matches made by adults from 54% in baseline to **96%** in the modified task.

The present study

The hypothesis of the present studies, drawn from our previous work, is that making shape and/or color unlikely bases of matching in a standard RMTS task (by varying shapes only on size) will allow four-year-old children, a population which ordinarily fails RMTS, to succeed *spontaneously*, that is *with no training or other preceding experience whatsoever*. Such a result would be important for at least two reasons.

1) Wholly spontaneous success on RMTS by four-year-olds is unprecedented

This would be the first *entirely spontaneous* success on the standard RMTS task in four-year-olds (i.e. with no training, demonstration trials or, in fact, *any* preceding content). Such a result would moreover provide further support for the conclusions of previous work (e.g. Walker & Gopnik, 2014; Kroupin & Carey, under review) that population differences between children of (at least) this age and older children/adults are differences in inductive biases *alone*.

2) Exploring the nature of inductive biases relevant to relational reasoning

Spontaneous success on RMTS without training as a result of simply changing dimensions on which stimuli vary would provide an important contribution to the ongoing discussion regarding the *nature* of inductive biases relevant to relational reasoning. In particular, there are two broad positions articulated in current research regarding the *scope over which* such biases are specified:

<u>1 - Biases with respect to all relations/object features (within a domain or across all domains)</u>

Most accounts which have discussed population differences in relational reasoning as differences in inductive biases have assumed that these biases lead individuals to preferentially attend to *all relations* or *all object features*. In some cases, this preference is specified within a particular context or "domain"- e.g. within the context of assessing cause-effect relationships (the “causal domain” - Goddu et al., 2020, see also Walker & Gopnik, 2014). Likewise, Gentner’s (Gentner, 1988, see also Gentner & Ratterman, 1991; Ratterman & Gentner, 1998) proposal that children have an early “object bias” within a given domain of reasoning presupposes that children attend to *all* object features over *all* relations within this domain. In other cases, the preference is proposed to be entirely domain-general: Carstensen et al. (2019) argue that the difference between Chinese three-year-olds who succeed on a simplified RMTS task and US three-year-olds, who do not, is a difference in a general “relational focus”. Similarly, Simms and Richland (2019) argue that priming four-year-olds with an analogy task makes them more likely to attend to relations in a “content-general” way by inducing a “generalized relational mindset”. In sum, these accounts propose that for a given domain (or even over *all* domains) inductive biases relevant to relational reasoning specify how likely *relations in general* are to be inferred as relevant relative *object features in general*.

2 - Biases with respect to particular representations

In our previous work (Kroupin & Carey, 2021a; Kroupin & Carey, 2021b; Kroupin & Carey, under review), we proposed that population differences in inductive biases across populations are specified over *particular representations*. All domains have an infinity of possible relational and object features, but for the purposes of illustration let us imagine a domain which has *only* relational features 1,2,3 (e.g. sameness, monotonic increase, equidistance from a given point etc.) and object features A,B,C, (e.g. color, shape, size, etc.). Previous accounts presuppose that inductive biases favor inferring as relevant *all of* [1,2,3] over *all of* [A,B,C] - or vice versa. In contrast, if inductive biases are specified over particular representations, it is perfectly possible for inductive biases to specify an ordering of relevance such that some relations are more relevant than some object features - but also some object features are more relevant than some relations (e.g. [1,A,2,3,B,C][16]). That is, for any given domain, an individual’s inductive biases can favor *some* relations over *some* object features and vice versa. This account leaves open the possibility of constellations of inductive biases in which all relations are more likely to be inferred as relevant than all object features (e.g. [1,2,3,A,B,C])

[16] In ordering these as a list I do not wish to make any claim about the format of representation. The point is merely that a precondition actually responding to the task *at all* is the presence of *some* mechanism by which a representation or subset of representations are selected as relevant bases of responding.

- but only as a particular case, not (*contra* previous accounts) as the way in which such biases are *necessarily* specified.

Discriminating between accounts

If the inductive biases of four-year-olds in our sample are specified over a given domain (or all domains) an in the generalized way, there is *no way in which* simply changing the dimensions on which RMTS stimuli vary should change children's performance since there is no meaningful sense in which the two the original and modified RMTS tasks constitute different domains. That is, if children's pre-existing inductive biases favor *all object features* within the domain in which RMTS tasks fall, then they will inevitably choose one of the infinity of possible object feature matches in the modified RMTS task (e.g. partial size matches) over matches on the relations same/different (recall that, unlike our toy 123/ABC example above, any really-existing domain has an infinity of such features). In contrast, our previous work predicts that decreasing the likelihood that shape and/or color *in particular* are inferred as correct bases of matching will increase the likelihood of matching by same/different relations *over any potential object feature matches*. In other words, an account of inductive biases specified over particular representations *uniquely* predicts that it should be possible to *produce spontaneous relational matching in a population which ordinarily fails to do so without changing pre-existing inductive biases in any way.*

In sum, the present work stands to make several important contributions to the literature on relational reasoning. Experiment 1 provides the relevant empirical data. First, however, I will briefly describe the data, drawn from our previous work (Kroupin & Carey, under review), from four-year-olds on an RMTS task varying in shape and color which will serve as baseline data.

Baseline RMTS

In our previous work, we tested four-year-olds drawn from the same population as Experiment 1, below, (mid- to high-SES, Boston families) on an RMTS task in which stimuli varied in shape and color (Figure 1) with no preceding training. The testing procedure and stimuli were identical to the generalization task of Experiment 1, below. Four-year-olds did not perform above chance on this version of RMTS without preceding training: Only one out of 24 children chose correctly on seven or eight out of eight RMTS trials and thus was above chance on a binomial test (p = .04). We will refer to children performing above chance as 'succeeders'. The proportion of succeeders in the sample was not statistically greater than chance on a second order binomial test (p = .35). As a group, children made fewer relational matches than would be expected by chance (41% v. 50% relational matches, p = .03, t = 2.24, Figure 4). Children's performance on a single triad was driving this unusual below-chance result (see Kroupin & Carey, under review). In order to be conservative in estimates of children's success, I will compare performance in Experiment 1 both to this baseline performance (41% relational matches) and to chance performance (50% relational matches).

Experiment 1

Experiment 1 tests the hypothesis that failure of four-year-olds on the standard RMTS paradigm, as in Experiment 1, is at least in part a result of pre-existing inductive biases which lead them to infer partial matches on shape and color to be the correct kind of match in RMTS, despite being *capable* of matching on the relations same and different. In order to test this a modified RMTS task was developed in which shape and color were equated across choice cards. Specifically, the standard RMTS task is modified such that the information about same-different relations is conveyed not by shape/color but by *size* (size-only RMTS, Figure 2).

Participants

Participants were 24 English-speaking children aged 49 to 60 months (M = 53.54m, 10 girls, 14 boys) recruited by phone from the greater Boston/Cambridge. Four additional children participated in the study but were excluded due to experimenter error (one child) or parental interference (three children).

Size-Only RMTS

Materials

Each RMTS trial contained three laminated paper cards, each of which contained two geometric objects (Figure 1). Unique objects were used on every trial. On each trial two choice cards were placed level with each other and below a sample card. One choice card was always 'same' and the other was always 'different'. On four trials the sample card was 'same' and on four trials it was 'different'. The correct choice card was the one which contained the same relation as the sample card. Sets of three cards were arranged into triads with one card as the sample card and two cards as the choice cards. The composition and arrangement of individual triads were the same across all participants, but the order in which the triads were presented was randomized - with the constraints that the task did not begin with more than two consecutive 'same' or 'different' trials and that there were never more than three consecutive 'same' or 'different' trials. The correct choice card was the one which shared a relation with the sample card (i.e. 'same' cards go with 'same' cards and vv. for 'different' cards).

Unlike in the RMTS task used in baseline data and in the generalization task below (Figure 1), all objects in this size-only RMTS task were black and all objects across both choice cards had the same shape. Objects on one choice card are the same size (instantiating the relation same), while objects on the other choice card are different sizes (instantiating the relation different). Objects on the sample card are the same shape as each other but of a different shape than the shapes on choice cards. Objects on sample cards in half of the trials are the same size (same trials) and on half of the trials are different sizes (different trials). The side on which the correct choice card was displayed and the side on which the smaller figure appeared on each 'different' card were counterbalanced. The composition and arrangement of individual triads of sample and choice cards were the same across all participants, but the order in which the triads were presented was randomized - with the constraints that the task did not begin with more than two consecutive 'same' or 'different' trials and that there were never more than three consecutive 'same' or 'different' trials.

It is critical to note that while size-only-O RMTS is a modification of the standard paradigm, *it is still a standard RMTS task*, that is, the abstract relational structure of the task is *identical* to the original: Success still depends on representing and aligning the relations same and different.

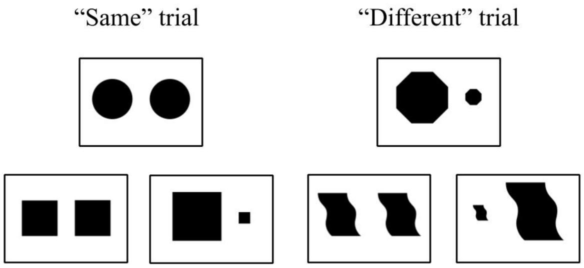


Figure 3. Two size-only RMTS trials

Procedure

Children were told that they would be playing a matching game. Choice cards were produced first and placed on the table as the experimenter said "Which one of these two cards…", then the sample card was produced and placed on the table as the experimenter finished the question "...goes with this card?" After children selected one of the two choice cards, the next trial was presented. No feedback of any kind was given during RMTS trials. After all eight trials were completed children were given a high five and a small prize for participating.

Potential non-relational strategies of responding in size-only RMTS

There are at least two non-relational bases of matching which children may use in size-only RMTS which would lead to successful performance on the task: First, children may match on overall symmetry on the card. However, previous work has shown that children do not match on whole-object symmetry until at least eight or nine years old (Shao & Gentner, 2019) making this strategy implausible in four-year-olds. If children match on the symmetry of the *set* of two objects on the card, this necessarily involves first identifying the two objects in the set as *the same* (or *different*) - equivalent to succeeding by simply matching on these relations directly.

Second, children may be making *iterated size matches*. That is, they may focus on objects in the sample card one at a time and search for an object of matching size in the sample cards. Doing so would lead them to match same cards with same cards not because of the shared relation, but because each had two medium-sized objects, and to match different cards with different cards because each had one large object and one small object. Given the very low rates of spontaneous matching on size by four-year-olds in simple Size MTS (Kroupin & Carey, under review) this strategy is unlikely. Nevertheless, here we addressed this possibility directly, with four extra trials.

Iterated-size matching control trials

Extra control trials always presented a sample card with two small objects of the same shape and color (black) *or* two large objects of the same shape and color, as well as one choice card with two medium objects and one choice card with a small object and a large object (Figure 4). If children were implementing an iterated size matching strategy to succeed on size-only RMTS they should match the sample card with the choice card containing a large and a small object since one of the two would constitute a match in size with one of the objects on the sample card. In contrast, if they were matching by the relations same and different, they should match the sample card to the choice card with two medium objects since both instantiate the relation same. Extra trials were administered with no break in procedure from size-only RMTS.

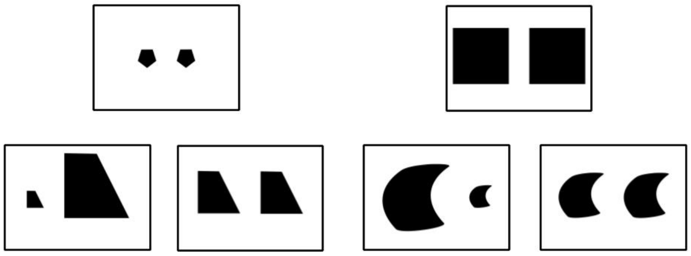


Figure 4. Two types of extra trials

Generalization to RMTS varying on shape and color

In order to see whether children continued matching on relations once the possibility of partial color and shape matches was re-introduced, the RMTS task with stimuli varying on shape and color from Kroupin & Carey (under review) was included after the completion of extra trials in Experiment 2. The hypothesis under consideration here makes no clear prediction in this case: It is possible that success on size-only RMTS (and extra trials) would change children's inductive biases so as to make them more likely to infer matches on sameness/difference as relevant in a subsequent task. However, it is also possible that the re-introduction of shape and color information would increase the salience of partial shape/color matches sufficiently to lead children to fall back to these as bases of matching (and thus not perform above chance, or at least markedly worse than in size-only RMTS). *Note once again* that there is no sense in which this latter task is *more* of a real RMTS test than the size-only version - both involve the *exact same* representational and computational requirements - children must match same to same and different to different. Materials and procedure were identical to the baseline task (Kroupin & Carey, under review)

Materials

The organization of same-different pairs within triads was identical to size-only RMTS, as was the number of trials (eight) and the number of trials featuring same pairs as the sample (four). Unlike size-only RMTS cards, however, figures on the generalization task had the same approximate height and width but varied in shape and color, such that no shape or color was repeated across cards (Figure 1)

Procedure

After completing extra trials, children were told that they would now be playing a new game, after which RMTS varying on shape and color was administered using exactly the same procedure as size-only RMTS.

Results

Size-only RMTS

11 out of 24 children chose correctly on seven or eight out of eight size-only RMTS trials and thus were above chance on a binomial test ($p = .04$). We will refer to children performing above chance as succeeders. The proportion of succeeders in the sample was statistically greater than chance on a second order binomial test ($p < .0001$). As a group, children made more relational matches on size-only RMTS than in the RMTS baseline task (72% v. 41%, $p < .001$, $t = 5.05$) as well as more than would be expected by chance (72% v. 50% relational matches, $p < .001$, $t = 4.28$).

Iterated-size control trials

We considered children who chose the relational match on the extra trials three or more out of four times to have consistently matched on relations, while those matching three or more out of four trials on object size to have consistently matched on size. Critically, if children succeeding on size-only RMTS were doing so by using iterated size matches, succeeders on size-only RMTS should consistently match on size in the extra trials. In contrast, if children were succeeding by matching on relations, succeeders on size-only RMTS should consistently match on relations in extra trials. The results were definitive: Of the 11 succeeders on size-only RMTS *none* consistently matched on size and *all 11* consistently matched on relations. This rules out the possibility that children's success on size-only RMTS was a result of a non-relational strategy of iterated size-matches. The remaining 13 participants averaged 2/4 matches on relations, with an equal number (four) of participants matching on relations 3-4 times as matching 0-1 times and the remaining five participants matching twice on each kind.

RMTS varying in shape and color

The criterion for succeeders on RMTS was the same as that for size-only RMTS (+7/8 trials correct). Nine out of 24 children were succeeders on the RMTS with stimuli varying in shape in color. As a group, four-year-olds performed better on RMTS than at baseline (Experiment 1 from Kroupin & Carey, under review; 69% v. 41% relational matches, p < .001, t = 4.14). The proportion of succeeders was also greater than in the baseline data (1/24 v. 9/24, Fischer's exact, p < .0001). This performance was also better than chance both in the overall proportion of relational matches made by four-year-olds (69% v. 50% relational matches, p < .001, t = 4.28, Figure X) and the proportion of children making seven or more relational matches (second-order binomial, p < .0001). There was no significant difference in performance between size-only RMTS and the subsequent RMTS task varying on shape and color (72% v. 69% relational matches, p = .63, t = .49).

Discussion

Simply changing the dimensions on which RMTS stimuli vary makes the difference between abject failure and robust success in four-year-old children. This result clearly confirms the initial hypothesis that making shape and/or color unlikely bases of matching in a standard RMTS task (by varying shapes only on size) will allow four-year-old children, a population which ordinarily fails RMTS, to succeed *spontaneously*, that is *with no training or other preceding experience whatsoever*. These results demonstrate *wholly spontaneous* success on standard RMTS in four-year-olds. These results also provide direct evidence regarding the *scope at which* inductive biases responsible for population differences in relational reasoning are specified. Furthermore, these results have implications for our theories of how inductive biases relevant to relational reasoning differ and change across phylogeny and ontogeny, for task interpretation and design for accounts of population differences based on inhibitory capacities. The following sections review these points in turn.

1) Unprecedented, spontaneous success

Previous work has shown that children age four or younger are capable of success on RMTS tasks, though exclusively in such cases when either 1) the structure of the task was simplified by using only one relation as the sample (e.g. Christie & Gentner, 2014; Walker & Gopnik, 2014) and/or 2) children were given experience with relational labels or relational matching prior to RMTS test (e.g. Kotovsky & Gentner, 1996; Christie & Gentner, 2014). In the former case, it is difficult to be sure that children are succeeding on the task by *matching* relations on every trial. The alternative is that children may identify which relation (same or different) constituted the correct choice, either as a result of demonstration trials, as in Walker and Gopnik (2014) or inferring the meaning of a relational label as in Christie and Gentner (2014) and simply choose the pair which instantiated this relation on every trial. Such strategies *do not involve* aligning relations and as such falls short of relational reasoning (see Kroupin & Carey, under review for a full discussion).

If those cases where children were given experience with relational labels or relational matching, it is possible that such experience *produced*, for the first time, representations of sameness and/or difference. For instance, Kotovsky and Gentner's (1996) progressive alignment paradigm trains children to match first on both object features *and* relations (e.g. AA goes with BC or AA), and then tests on relational matching alone (e.g. DD goes with EE or FG). Kotovsky and Gentner (1996, see also Gentner & Hoyos, 2017) argue that this procedure allows children to develop new 'relational abstractions', i.e. representations of relations (here of sameness and difference) in a new format. If training stands to have this kind of effect, it remains possible that children's initial failures were *not* a result of differences in inductive biases alone, but rather as a result of lacking these representations coming into the task.

In contrast our previous work (Kroupin & Carey, under review) and, even more emphatically, Experiment 1 above, demonstrate conclusively that, for at least a large portion of four-year-olds (11/24 in the present sample) the population differences in RMTS performance is due to differences in *inductive biases alone*. Needless to say this *by means rules out* population differences in representational or computational capacities (e.g. *per* Penn et al., 2008; Halford, 1993, respectively), or the presence/absence of the necessary representations (e.g. Christie & Gentner, 2014) - even in the case of the approximately half of four-year-olds in our sample who did not succeed on RMTS in Experiment 2. The crucial point, however, is that unless otherwise demonstrated any given population difference is *potentially* a difference in inductive biases relevant to relational reasoning - the nature of which this work also begins to shed some light on. Again, you can't introduce these distinct ways of thinking about population differences as if you've already explained what you are talking about. "The curse of knowledge" rears its ugly hea

2) The nature of inductive biases responsible for population differences

The present studies

The dramatic difference in spontaneous performance (abject failure v. robust success) across two RMTS tasks depending entirely on what dimensions the stimuli vary cannot be explained by accounts which assume that inductive biases specify a *general* preference for object features or relations across a domain (or *a fortiori*, across all domains). After all, there is no meaningful sense in which the two RMTS tasks are in different *domains*, yet clearly the inductive biases of four-year-olds in our sample lead them to infer object features (partial shape/color matches) as correct bases of matching in one case, and relational features (same/different) in the other.

An alternative account of inductive biases involved in population differences in relational reasoning posits that inductive biases underpinning population differences in relational reasoning performance are specified over *particular representations*. Such an account readily explains the present results: If four-year-olds' inductive biases in RMTS are organized, in order of likelihood of inferring as a correct basis of matching, as [partial shape/color, sameness/difference, partial size] they will fail the standard task by virtue of attempting partial shape/color matches (e.g. ‘these two go together because they’re both green-ish’). In this case, simply reducing the likelihood of partial shape/color matches being inferred as correct should increase the likelihood of same/different matches *without any changes to pre-existing inductive biases* - as is demonstrated by the results of Experiment 2.

Domain-wide/Content-general biases

Of course the fact that there *are* population differences in inductive biases specified over particular representations in no way rules out differences at content-general or domain-wide levels. Returning to the example above of a domain with relational features 1,2,3 and object features A,B,C - a particular-representation account which allows for an ordering like [1,2,A,B,3,C] *also allows for* orderings like [1,2,3,A,B,C] or [A,B,C,1,2,3]. The crucial point is that domain-wide (or domain-general) accounts *cannot* accommodate the former kind of ordering - which Experiment 2 provides direct evidence of.

The same logic applies to effects of training tasks on inductive biases: Simms and Richland (2019) argue that priming children with distant analogies (i.e. having them complete, with a parent's assistance, analogies like caterpillar:butterfly::egg:?) leads to a 'content-general' attention to all relations. Specifically, they demonstrate that priming with distant analogies leads to an increased likelihood of matching according to relations (e.g. which individual participates in the chasing relation as the chaser in a scene) in a task where object feature matches are also available. That is, using the formalism above, Simms and Richland (2019) argue that priming by distant analogies changes *any previous ordering* of inductive biases into [1,2,3,A,B,C] such that 1-3 represent *all relational representations* (including 'the chaser') and A-C represent *all representations of object features*[17]. However, given the infinity of possible relational and object feature representations available for any given stimulus, this kind of change would be computationally intractable. The same empirical result can, however, be explained by distant analogies affecting the inductive biases relevant to a *particular* representation -e.g. making

[17] Vendetti et al. (2014) make the same argument using similar tasks with adults.

(partial) shape matches *less likely to be inferred as correct* since the corresponding analogues do not closely correspond in shape (e.g. butterfly, chicken). Thus this priming may in fact change the ordering [shape, chaser-status] to [~~shape~~, chaser-status] (no doubt there are infinitely many other features and relations in this ordering - here I highlight only the relative change in the two most relevant ones).

A more generous interpretation of Simms and Richland's (2019) account of their priming effects (*pace* their explicit description of a "generalized relational mindset") can avoid positing a computationally-intractable process of updating *all* relational features relative to *all* object features - albeit at the cost of posting several additional mechanisms. Namely, it is possible that inductive biases lead individuals to select *not* a single feature most likely to be relevant to the task (in RMTS, a single basis of matching e.g. partial shape matches) but rather a *subset* of such features in ranked order. *If* this is the case, *then* it is possible that the effect of analogical priming is that, after priming has occurred, any ordered subset which is generated by pre-existing inductive biases is adjusted such that all relations become relatively more likely to be inferred as relevant. Schematically, out of a domain which contains relational features 1,2,3 and object features A,B,C, an individual's inductive biases may lead them to infer [A,1,C,3] as an ordered subset of features likely to be relevant to responding to the task (e.g. adults in Kroupin & Carey, 2021 may infer [matches on sameness, incomplete object matches], see Figure 3). The effect of analogical priming may then be to increase *in general* the relative salience of relational features *within this subset*, changing it to e.g. [1,A,3,C] or [1,3,A,C]. In this sense, training tasks could *generally* increase attention to relations without any computationally intractable operations. However this 1) requires machinery for which we, at present, have no evidence and, more importantly, 2) *presupposes* inductive biases specified over particular representations since in order to form the necessary subsets *in the first place*.

Biases over the causal domain

The proposal that children's inductive biases make them more likely to attend to relational features in a causal domain (e.g. Walker & Gopnik, 2014; Goddu et al., 2020) is intriguing and supported by clear empirical results. That being said, it is not clear *why* this should be the case. Goddu et al. (2020) suggest that one possibility that causal reasoning is inherently relational (i.e. necessarily involves the relation *cause*). However, the fact that an operation, e.g. causal reasoning, which instantiates relations in no way implies that it involves *representations* of relations - the latter of which are required for second-order relational reasoning. For instance, solving a simple match-to-sample (MTS) task (e.g. Does A go with A or B?) involves identifying which two objects present in the trial are the same - with sameness obviously being a relation. Yet, success on MTS *need not involve a representation of a relation*. For instance, MTS can be completed using an operation like [Store X in working memory, Seek X], where X is some feature of the stimulus - an operation which *instantiates* yet *contains no representation of* the relation same[18]. Occam's Razor suggests that, generally, an operation *instantiating* a relation will not contain a *representation* of this relation *unless the latter is explicitly required*. Whether this is the case in the case of causal reasoning is, of course, an empirical question.

However, even if causal reasoning inherently increases attention to relations, the mechanism by which it does so faces the same issues as content-general increases in attention to relations: A causal context cannot increase attention to *all* of the infinity of relations within the

[18] See Hochmann et al., 2016; Kroupin & Carey, under review; under review and Zentall et al., 2018 for evidence that this is, in fact, how MTS is solved by infants, children, adults and pigeons, respectively

stimuli. If it increases attention to *some set* which is initially attended to then this set still must be defined in terms of inductive biases over particular representations, as discussed above.

3) Implications for change over phylogeny and ontogeny

If differences in inductive biases relevant to relational reasoning are specified at the level of particular representations, it follows that *changes* to these biases will also occur at least at this level (though more general changes are possible - as discussed above). It follows that the particular pattern of inductive biases in any individual or population will depend heavily on their *particular* set of experiences (in addition to potential relevant genetic differences). In humans, this means that the inductive biases relevant to relational reasoning will likely vary significantly by culture, since culture significantly determines the kind of stimuli children experience and the representations which are salient in their day-to-day lives. Cross-cultural work has, in fact, shown a great deal of variability in the degree of attention to relations across cultures - with differences between collectivist (Eastern) and individualist (Western) cultures a particular focus of study (e.g. Carstensen et al., 2019, Imada et al., 2013; Kuwabara & Smith, 2012).

In non-human animals, inductive biases will also likely depend in large part on their environment - a natural environment in the case of wild animals, or a lab environment in the case of animals raised and/or kept in such. The latter crucially includes previous experience with various stimuli and tasks - some of which may produce inductive biases directly relevant to performance on tasks such as RMTS. For instance, training on Identity MTS may produce inductive biases relevant to matching tasks with geometric figures such that shape/color/size are inferred as the correct bases of matching. This may *interfere with* learning a subsequent RMTS task if the initial biases lead participants to consistently attempt feature matches and ignore relational ones. The possibility is not a theoretical one - *at least some* comparative studies (e.g.

Fagot, Wasserman & Young, 2001) and potentially *almost all* (according to a reviewer on Kroupin & Carey, under review) involve training on Identity (or at least Shape) MTS prior to testing on RMTS.

4) Implications for task design and interpretation

Interpretation of performance depends upon understanding pre-existing inductive biases

The issue of which inductive biases exist in a population prior to testing leads us neatly to the implications of the current results for the design of tasks like RMTS and interpretation of performance thereon. Clearly, we *cannot assume* that participants' pre-existing inductive biases will be neutral with regard to their performance on a given task - this is shown quite plainly in the contrast between the two RMTS tasks in Experiment 1 (size-only and varying in shape and color). It follows that in designing tasks and interpreting their results, we need to have some assessment of the pre-existing inductive biases of the population we are testing. In the case of RMTS, this includes children's predisposition to match on shape and/or color - as well as the fact that they do not seem to have a bias against making *partial* (approximate) matches on either of these two dimensions (e.g. 'these two go together because both have a shape that is pointy'). It seems likely that the same principle will hold in work with non-human animals. Namely, that their performance will vary significantly depending on previous experience with tasks using similar stimuli - especially if these tasks also involved matching.

Inductive biases are not just in the individual - they are built into the task

Unlike previous paradigms, discussed above, which lead children to succeed by *changing* their inductive biases, the current work takes the opposite approach and changes the *task* so as to align with children's pre-existing inductive biases. The possibility of this strategy highlights an obvious but critically important feature of studying inductive biases relevant to population differences: Performance is determined not simply by an individual's inductive biases, but how these biases *correspond* with the inductive biases *presupposed by the task*. That is, determining any action as relevant depends upon solving a vast number of inductive problems, and, consequently, meeting any metric of performance presupposes a (or a set of possible) particular constellation(s) of inductive solutions. RMTS provides a case in point: There is no *objective* sense in which matching by sameness/difference is more correct than e.g. making partial shape matches. Consequently, success requires a *particular* inference - one which participants will make given a *particular* set of stimuli if they have a *particular* set of inductive biases. This latter set is the one which is *presupposed by* or, in other words, *built in to* the RMTS task.

It follows that our approach to designing tasks depends directly on what we are attempting to study: If our goal is to study whether a certain cognitive capacity exists *at all* in a population, we must take pains in making sure that our task fits the pre-existing inductive biases of the populations so as to make it maximally likely that they will infer the capacity in question - and not some other strategy - as relevant in solving the task. And, to the extent that, as in the case of US four-year-olds and RMTS, the relevant inductive biases are specified over *particular representations*, this fitting process must be relatively detailed e.g. including adjusting the dimensions on which stimuli vary to individual's inductive biases.

If, on the other hand, our goal is to study the likelihood with which an individual will use a particular cognitive capacity in *real-world situations*, then we must ensure that the inductive biases built into the task maximally resemble the real-world situations we are interested in (e.g. entrance exams, particular professional problems etc.)

5) Interaction with changes in inhibitory capacities

The success of four-year-olds in Experiment 2 rules out the possibility that populations which initially fail on RMTS fail due to a lack of representational capacity. Likewise it rules out, in this case, accounts of population differences in terms of working memory capacity (e.g. Halford, 1993) since the working memory requirements of the RMTS tasks in Experiments 1 and 2 do not differ. Another account, however, proposes that populations differ in relational reasoning performance as a result of differences in *inhibitory* capacity (Richland, Morrison & Holyoak, 2006). That is, children fail to match on relations because they cannot inhibit prepotent responses on object features.

At first glance, such an inhibition-based account may be compatible with the data presented here: If children are disposed to attend to shape/color matches, the presence of partial shape/color matches in the RMTS of Experiment 1 may tax their inhibitory capacities more significantly than the RMTS of Experiment 2. Such an account cannot, however, explain the fact that, in Experiment 2, performance on standard RMTS did not significantly differ from performance on the preceding size-only RMTS task. After all, once shape/color variation is reintroduced into the stimuli, an inhibition-based account predicts sharp decline in performance.

More importantly, however, an inhibitory account of the present results *presupposes* that children *were trying to* inhibit attention to partial shape/color matches in standard RMTS and *failed to do so*. This is far from obviously the case. A simpler explanation is that children simply inferred partial shape/color matches as *more relevant* than relational matches (if they noticed the latter at all). More generally, the inhibitory difficulty of a given task is (at least in part) determined by the aspects of the stimuli the participant's inductive biases *spontaneously lead them to attend to*. As a result, any account of inhibitory changes in relational reasoning across ontogeny or phylogeny is *preconditioned on* 1) an account of inductive biases of the populations being studied *relative to* the inductive biases built into the current situation such that 2) two sets of features are inferred as relevant, one of which is more salient and needs to be inhibited in order to attend to the one which is inferred to as, ultimately, most relevant.

Conclusion

While the present results provide clear evidence that population differences are *sometimes* differences in inductive biases specified over particular representations, it would be foolish to argue that this is *always* the case. Indeed, more than half of four-year-olds do not succeed even in the size-only version of RMTS, leaving open the possibility that at least some children at this age genuinely *cannot* succeed on the task (for lack of the necessary representations or computational capacities).

That being said, had we taken the results of the baseline RMTS task at face value, we may have said the same for *all* four-year-olds (*pace* Kroupin & Carey, under review). The fact that this is not the case underscores the importance of integrating accounts of inductive biases into our theories of relational reasoning - *both* as a necessary reference point for interpreting our results *and* as an independent component of the development of relational reasoning. After all, even if some mythical child is born with unlimited representational and computational capacities, her actual *performance* on RMTS, or any other experimental or real-world task involving relational reasoning, will be strictly limited by the degree to which her inductive biases guide her to choose *which* of the infinity of possible relational comparisons available is the correct one to apply in the task. In sum, whether we are interested in the *capacity* of a given population to engage in relational reasoning or in the *proficiency* with which they deploy relational reasoning in day-to-day situations, we must characterize the relevant inductive biases - those of the individuals coming to the task and how well they fit with the inductive biases built into the task.

CHAPTER 5

Conclusion

Population differences in relational reasoning are sometimes differences in inductive biases alone

Papers 1-3, and the latter two in particular, established for the first time that population differences in full-blown relational reasoning are, in some cases, differences in inductive biases formulated over particular representations *alone*. Specifically, around half (11/24 in Paper 3), if not more of four-year-olds in a Western, middle-class sample initially fail RMTS as a result of inferring partial matches on shape/color as the correct bases of matching *despite being capable of* spontaneously matching on the relations same and difference (Paper 3). Previous accounts of population differences on RMTS presupposed that those populations that did not spontaneously succeed on the task lacked the necessary representations and/or computational mechanisms. Were this the case, unitary representations of sameness and difference would be *unique*, in the natural world, to human children over the age of five (in Western samples at least). The fact that at least in some cases lack of spontaneous success (and even persistent failure in the face of training) is a result of differences in inductive biases *alone* means that such representations are more widely available across ontogeny, and potentially phylogeny *without training*.

These results have at least three important theoretical implications - two negative and one positive: First, this means that we cannot rule out that such representations feature as primitives in an innate language of thought. Second, the uniqueness of human adult relational reasoning is *not exclusively* a function of a unique format of representation which emerges in late childhood (as suggested by Penn et al., 2008). Third, the unique pattern of relational reasoning performance seen in human adults *depends upon* a constellation of inductive biases which develop across age and, almost certainly, vary across culture. The first of these implications further motivates an ongoing research program focusing on determining whether unitary representations of sameness and difference are available innately whether in humans or non-human animals (e.g. Hochmann et al., 2016; Martinho & Kacelnik, 2016). The second two implications motivate a number of considerations including how we think about the role of inductive biases in our methods and in our theories of human-uniquene relational reasoning, the way in which we may think of RMTS as an ecologically (in)valid task, and motivate at least one of two systematic research programs into the nature and development of inductive biases relevant to relational reasoning across ontogeny and phylogeny. The following sections address each of these in turn.

Inductive biases in the development of relational reasoning

Inductive biases in task design and interpretation

The results in Papers 1,2 and, especially Paper 3 highlight the importance of developing accounts of inductive biases for the populations we test using tasks like RMTS. After all, as emphasized in Paper 3, we simply *cannot* interpret failure on such tasks without appealing to a baseline of pre-existing inductive biases, and how these interface with the inductive biases presupposed by (built in to) the task design. Future work attempting to establish which populations do/do not possess unitary, symbolic representations of same/different must work to fit their tasks (RMTS or otherwise) to the pre-existing inductive biases - or at least attempt to instill the relevant inductive biases in participants *without* producing new relational representations (as with the MTS training in Paper 2). Failure to do so will lead to the impossibility of distinguishing failures due to a lack of the relevant representations from failures due to inductive biases alone.

Returning to RMTS specifically, if culture is one of the (though certainly not the only) main determinants of inductive biases, and RMTS performance depends on having the right inductive biases, it is reasonable to expect cultural variation on the task. Unpublished data from

Pitt and Piantadosi provide exactly such evidence: Working with the Tsimane, an indiginous group in the Bolivian Amazon with limited access to formal education, Pitt and Piantadosi found that *adults* in this population struggled to succeed on RMTS, despite being given repeated demonstration trials and even instructions. In a very real sense, then, the difference we observe between (at least some) four-year-olds and adults in our population on RMTS is a *cultural* difference - with the former not having yet had the culturally-specific learning experiences necessary to produce the inductive biases which lead US adults to spontaneously succeed on the task.

Human uniqueness and inductive biases

The results in Papers 1-3 draw our attention to a point in the pattern of RMTS performance across population that holds for relational reasoning across populations generally: Setting aside any differences in *representations* and *computations* - the cultures in which adult humans participate *expect* and *provide opportunities for* vastly more relational processing than the environments of non-human animals and even young children. Our inductive biases, consequently, must develop to keep pace with these expectations and opportunities, such that we are ready and able to deploy our cognitive capacities to successfully navigate our environments. For instance, *only* adults and children over a certain age are expected and have the opportunity to learn and tell stories featuring metaphors, to (in some cultures) use mathematics to model the world. Consequently, by humans in the relevant cultures must, by adulthood, have developed the appropriate inductive biases to interpret and make use of these situations. Just the same, *only* adults *in certain cultures* are expected and (typically) have the opportunity to engage in the kind of syllogistic reasoning problems studied, Luria, Harris and their colleagues. Adults in these cultures (with formal education) develop the inductive biases that lead them to spontaneously succeed on syllogism tasks, while children from the same culture have not yet done so - and individuals from other cultures never do.

These expectations/opportunities do not arise *simply as a function of* possessing the necessary capacities to engage in these practices *nor* from having developed the necessary representations. Nor, even, are they a result of changes across ontogeny or phylogeny in domain-general attention to relations (as posited by Vendetti et al., 2014; Simms & Richland, 2014) - as evidenced by Paper 3 and the data like those of Pitt and Piantadosi. Such patterns necessarily also involve complex systems of inference produced by cultural evolution (e.g. Boyd & Richardson, 1985) and internalized as norms of behavior and opportunities to act (in this case engage in relational reasoning) with the goal of self-expression and increasing wellbeing (e.g. Bourdieu, 1977; 1980). In other words, we are unique in the natural world not just in what we *can do or think in principle*, but in the *opportunities for doing and thinking* we create and see in the world[19]. Inductive biases are a critical part of what allows us to successfully navigate these opportunities.

Returning to ecological validity: What can RMTS tell us?

The fact that RMTS performance can be seen to vary dramatically as a result of inductive biases alone in some senses confirms the suspicion, referred to all the way back in the general introduction, that the task is problematic due to its lack of ecological validity. After all, the expectations and opportunities in the environments of neither four-year-olds nor, certainly, non-human animals are likely to feature matching geometric relations according to

[19] The points in this section and many similar ones have been developed at great depth by an an enormous range of authors, from Aristotle (ca. 350 B.C.E./1925), to Marx (1867/1992), the Soviet school of psychology (Vygotsky, 1929, Luria, 1971) and their Western followers (e.g. Cole, 1998; Rogoff, 2003) as well as researchers in the traditions of ecological psychology (e.g. Gibson, 1983; Barker, 1968) etc. Integrating their work here is far outside the scope of this dissertation. However, it would be remiss of me not to refer to the important role their works have had on my thinking about these issues.

same/different relations. Nevertheless, there are two reasons why RMTS remains of interest to research:

First, the fact remains that performance on RMTS is some of the strongest evidence we have for the possession of same/different representations in a unitary, symbolic format. Moreover, unlike having learned a symbol for sameness (e.g. "same" or plastic tokens with the same meaning taught to primates, Premack, 1983; Thompson et al., 1997), performance on RMTS can be assessed *without training* on relations - as we saw in Papers 2 and 3 - allowing us to examine which populations do/do not have such representations *without training*. As such, even if we have to make significant adjustments in the format of the task and the preceding training tasks so as to align the inductive biases of children/non-human animals with the task, RMTS testing remains an important tool in investigating the phylogeny and ontogeny of such representations. For instance, an important future step would be to test the possibility that MTS training drawn from the Smirnova et al. (2015) paradigm can produce spontaneous success on RMTS in non-human animals *without progressive alignment*.

Second, from the perspective of inductive biases, the pattern of spontaneous RMTS success across populations is important evidence of culturally-induced inductive biases. After all, if we grasp the enormity of the inductive problem posed by RMTS, the fact that the vast majority of US participants over the age of five *spontaneously match by same/different relations* gives us an important indication of the kind of inductive bias which living in this culture produces. That is, something about the expectations and opportunities of our culture means the by the age of five inductive biases which lead children to spontaneously infer comparing abstract relations among sets of two-dimensional shapes as a viable approach to problem solving. Clearly, these biases would not develop *for no reason*. Consequently, the fact that children in our culture *do* succeed spontaneously on RMTS indicates that there is *some* ecological validity of the inductive biases tapped by this task. In the same way, the fact that adults in Western countries *do* succeed on syllogism tasks indicates something meaningful about the inductive biases *of that group*.

The work of Vendetti, Wu and Holyoak (2014) is provides a tantalizing clue in this regard: Vendetti et al. (2014) found that those adults who spontaneously matched on relations in a task where both relational and object matches were available performed better on a test of Raven's matrices than those who spontaneously matched on objects. Raven's matrices performance, in turn, has been shown to correlate a wide variety of life outcomes, including job success and health (see Deary, 2012 for an overview). Crucially, once adults were primed with distant analogies (e.g. completing wood:woodstove::stomach:?) they became significantly more likely overall to spontaneously match on relations *and the group difference in Raven's matrices score disappeared*. That is, adults *pre-existing* inductive biases predicted Raven's performance but *not* once these biases had been altered by training.

Needless to say this is very limited evidence for the relationship between spontaneous relational matching and IQ test performance (much less life outcomes). Nevertheless, it illustrates the potential fruitfulness of research into the nature of the inductive biases tapped by spontaneous relational matching - as in the case of RMTS. In sum, RMTS clearly does not have ecological validity as a test of relational reasoning *in general*. But as a test of the contexts in which individuals spontaneously infer relational comparisons to be relevant, it may well have some ecological validity within the WEIRD societies from which our samples are overwhelmingly drawn. For instance, the age at which children spontaneously succeed on RMTS may well predict later outcomes e.g. in school, in the same way that children's spontaneous attention to the numerosity of sets robustly predicts mathematics performance over and above mathematical knowledge (see MacMullen, Chan et al., 2019; MacMullen, Vershaffel & Hanula-Sormunen, 2020 for recent reviews).

Future directions: RMTS

The clearest outstanding question in the RMTS literature to date is whether individuals without previous training in symbols for sameness/difference can succeed on RMTS. The present work does not clearly resolve this issue since four-year-olds in our sample generally know the words "same" and "different" (Hochmann et al., under review). Likewise, progressive alignment training included in the original Smirnova et al./Obozova et al. (2015) paradigm may have, at least in principle, led birds to form such a symbolic representation *de novo*. The same is true of dogged training paradigms (e.g. Fagot & Thompson, 2011, Truppa et al., 2011) and, most obviously, symbol training paradigms (e.g. Premack, 1983; Thompson et al., 1997). The most immediate next step in this line of work would be to replicate the experiments of Papers 2 and 3 in younger children (while also testing for same-different vocabulary knowledge) as well as in non-human animals (e.g. replicating the original Smirnova et al./Obozova et al. paradigm without progressive alignment trials).

Given the demonstrated importance of inductive biases on performance, and the importance of culture for determining inductive biases in humans, another important line of research on RMTS is systematic comparison across cultural groups. As discussed above, Pitt and colleagues have produced highly intriguing findings with a traditional-culture group (the Tsimane). Identifying the *nature* of cultural differences responsible for differences in performance will prove important for our understanding of the kind of information children in Western samples are exposed to which allows them, by the age of five, to succeed on RMTS spontaneously. In a similar vein, testing across SES levels *within* Western samples would provide suggestive evidence as to the necessary cultural input for spontaneous success is related to e.g. parental education. Finally, testing whether spontaneous relational matching predicts cognitive measures (e.g. IQ, as in Vendetti et al., 2014) or outcomes (e.g. school performance) may provide insight into the *real-world importance* of the differences in cultural input reflected in (the age of) RMTS success.

Future directions: Weak and strong research programs on inductive biases

The weak program of research into inductive biases

As discussed in this conclusion and in all of Papers 1-3, we simply *cannot* avoid characterizing the pre-existing inductive biases of the populations we study if we wish to avoid mischaracterizing failures due to differences in such biases as failures resulting from a lack of necessary representations or capacities. The weak program of research into inductive biases is simply one which works to specify these biases to the extent necessary for us to interpret task performance.

The weak program of research proposed here aims to produce *systematic data* on what inferences a typical pattern of inductive biases within a given population leads individuals from said population to make given certain kinds of stimuli/tasks (e.g. colored geometric shapes, puppets). This itself is not a trivial challenge - it is important to identify which tasks and stimuli sets are worth investigating (e.g. by virtue of being commonly used within one or more research programs) and explicitly testing hypotheses as to what *spontaneous inferences* different populations may have to these stimuli - and how these may differ from the inferences necessary to succeed on the task. Yet, the studies presented here have demonstrated on the case study of RMTS that meaningful progress can be made on such issues - with important consequences for our understanding of species-universal cognitive capacities for relational reasoning.

The strong program of research into inductive biases

The strong program of research into inductive biases extends the weak program by assuming that the constellation of inductive biases within a given population is not only *possible* but also *important to* study in and of itself. That is, the project involves describing the interrelation of individuals within a given population and their typical environments so as to build up a systematic picture of the inductive biases the former are likely to have and the latter are likely to produce. Such a systematic picture would, consequently, allow us to generate empirical hypotheses regarding the spontaneous inferences of this population in (at least some) novel contexts, not only facilitating our interpretation of performance on novel tasks but furthermore allowing us to better understand how the cognitive capacities available to these individuals are *actually deployed*. In the long run this involves performing what Cole and his colleagues (1973) described as 'cognitive ethnographies' - partnering with sociologists and anthropologists to detail the cognitive (and particularly inferential) expectations and opportunities of various environments, e.g. middle-class pre-schools, traditional-culture marketplaces etc.

There are many possible objections to the feasibility of the strong program and here is not the place to present them and my rebuttals. Fortunately, even if we reject the possibility of the strong program, we have no alternative but to continue with the weak program. And since the investigations involved in the latter are a subset of those involved in the former, we may make progress on this program of research while suspending any ultimate judgment on the strong program.

Final thoughts

The work in this dissertation has demonstrated that population differences in RMTS, and thus relational reasoning generally, are sometimes the result of differences in inductive biases formulated over specific representations. This result contributes to our understanding of same/different representations and the origin of relational reasoning across phylogeny and ontogeny. A great deal of productive future research can be and is being done on both of those fronts. In this conclusion, however, I have emphasized the implications of this work for a study of inductive biases - not because this line of work is unusually important *a priori*, but rather because it is (at least in the relational reasoning literature) markedly less well-developed than work on representational and computational issues.

Yet, there seems to me no alternative to moving down this path if we accept - as Papers 1-3 demonstrate that we must - the role of inductive biases in determining population differences in relational reasoning. If the differences are real, they will affect our results. The notion that we can produce *no* systematic characterization of population differences in inductive biases is ruled out by the studies above - and a great deal of, often cross-cultural, literature besides. If we *can* produce such systematic characterizations, we need to do so at least in order to understand differences between populations in responding to a particular task. And if we *do* start working on systematic characterizations, we will eventually have to decide whether these can and should be extended to more general treatments of our physical and cultural environments. In closing, I propose that we can do so, that we should do so and that doing so will make our studies of relational reasoning, and cognitive science more generally, richer and better able to reflect the meaningful diversity of thought across species, age, and culture.

References

Aristotle, ., Ross, W. D., & Brown, L. (2009). *The Nicomachean ethics*. Oxford: Oxford University Press.

Barker, R. G. (1968). *Ecological Psychology: Concepts and Methods for Studying the Environment of Human Behavior.* Stanford: Stanford University Press.

Bermúdez, J. L. (2003). *Thinking without Words*. Oxford: The Oxford University Press.

Bohr, N. (1913). On the Constitution of Atoms and Molecules. *Philosophical Magazine*, *26*(6), 1–25.

Bourdieu, P. (1977). *Outline of a Theory of Practice (Cambridge Studies in Social and Cultural Anthropology)* (Reprint ed.). Cambridge University Press.

Bourdieu, P. (1992). *The Logic of Practice* (1st ed.). Stanford University Press.

Boyd, R., & Richerson, P. J. (1988). *Culture and the Evolutionary Process* (First Paperback Edition). University of Chicago Press.

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.

Castro, L., & Wasserman, E. A. (2016). Attentional shifts in categorization learning: Perseveration but not learned irrelevance. *Behavioural processes, 123*, 63-73.

Carroll, L. (1865). *Alice's Adventures in Wonderland*. Macmillan Publishers.

Carstensen, A.B., Zhang, J., Heyman, G.D., Fu, G., Lee, K., & Walker, C.M. (2019). Early diversity in abstract thought: Context shapes the developmental trajectory of relational reasoning. *Proceedings of the National Academy of Sciences, 116*(28).

Chan, J.Y.C., & Mazzocco, M. M. M. (2017). Competing features influence children's attention to number. *Journal of Experimental Child Psychology, 156,* 62-81.

Chan, J. Y.C., Praus-Singh, T., & Mazzocco, M.M.M. (2020). Parents' and young children's attention to mathematical features varies across play materials. *Early Childhood Research Quarterly,* 50, 65-77.

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

Christie, S. & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11*(3). 356-373.

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*(2), 383-397.

Cole, M. (1990). Cognitive development and formal schooling: The evidence from cross-cultural research. In L. C. Moll (Ed.), *Vygotsky and education* (pp. 89-110). Cambridge, England: Cambridge University Press.

Cole, M. (1998). *Cultural Psychology: A Once and Future Discipline*. The Belknap Press.

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

Cook, R. G., & Wasserman, E. A. (2007). Learning and transfer of relational matching-to-sample by pigeons. *Psychonomic Bulletin & Review, 14*, 1107–1114

Darwin, C. (1871) *The descent of man, and selection in relation to sex*. John Murray

Davidson, D. (1982) Rational animals. *Dialectica 36*, 317–327

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*(1), 453–482.

Descartes, R. (1637/1985) Discourse on the method. In *Descartes: Selected Philosophical Writings* (Cottingham, J. et al. eds), pp. 20–56, Cambridge University Press

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168.

Fagot, J., & Parron, C. (2010). Relational matching in baboons (Papio papio) with reduced grouping requirements. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(2), 184–193.

Dias, M., Roazzi, A., & Harris, P. L. (2005). Reasoning From Unfamiliar Premises: A Study With Unschooled Adults. *Psychological Science*, *16*(7), 550–554.

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.

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.

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*, 316–328

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

Fodor, Jerry A. (1975) *The language of thought.* New York: Crowell

Fodor, J. A. (2010). *LOT 2: The Language of Thought Revisited* (Reprint ed.). Oxford University Press.

Gentner, D. (1988) . Metaphor as structure mapping: The relational shift . *Child Development, 59*, 47-59.

Gentner, D., & Forbus, K. D. (1991). MAC/FAC: A model of similarity-based access and mapping. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, 504-509.

Gentner, D. & Hoyos, C. (2017). Analogy & abstraction. *Topics in Cognitive Science, 9*(3), 672-693.

Gentner, D., & Rattermann, M. (1991). Language and the career of similarity. In S.A. Gelman & J.P. Bymes (Eds.), *Perspectives on thought and language: Interrelations in development.* London: Cambridge University Press

Gibson, J. (1983). *The Senses Considered as Perceptual Systems* (Revised ed.). Praeger.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355.

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.

Goddu, M. K., Lombrozo, T., & Gopnik, A. (2020). Transformations and transfer: Preschool children understand abstract relations and reason analogically in a causal task. *Child development*, *91*(6), 1898-1915.

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

Grant, B. (2007). Do chimps have culture? *The Scientist, 21*(8), 29-35. Retrieved from http://search.proquest.com.ezp-prod1.hul.harvard.edu/trade-journals/do-chimps-have-culture/docview/200056739/se-2?accountid=11311

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

Harris, P. L. (2001). Thinking about the unknown. *Trends in Cognitive Sciences*, *5*(11), 494-498.

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action‐based verb extension. *Child Development*, *82*(2), 674-686.

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.

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

Holyoak, K. J., & Thagard, P. R. (1989). A computational model of analogical problem solving. *Similarity and analogical reasoning*, *242266*.

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

Imada, T., Carlson, S. M., & Itakura, S. (2013). East-West cultural differences in context-sensitivity are evident in early childhood. *Developmental Science, 16*, 198–208

James, W. (1890). *The Principles of Psychology.* New York, NY: H. Holt and Company.

Jamrozik, A., & Gentner, D. (2020). Relational labeling unlocks inert knowledge. *Cognition*, 196, 104-146.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*(6), 2797-2822.

Kroupin I., Carey, S. (in press). Inference in relational reasoning: Relational-match-to-sample as a case study. *Journal Experimental Psychology: General*.

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.

Kroupin I., Carey, S. *You cannot find what you are not looking for: Population differences in relational reasoning are sometimes differences in inductive biases alone*. Manuscript under review

Kuwabara, M., & Smith, L. B. (2012). Cross cultural differences in cognitive development: Attention to relations and objects. *Journal of Experimental Child Psychology, 113*, 20–35.

Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, *31*(6), 807-825.

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

Long, B. & Konkle, T. (2017). A familiar-size Stroop Effect in the absence of basic-level recognition. *Cognition, 168*, 234-242.

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.

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Cambridge, MA: Harvard University Press.

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., … Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America, 111*(20), E2140–8.

McMullen, J., Verschaffel, L., & Hannula-Sormunen, M. M. (2020). Spontaneous mathematical focusing tendencies in mathematical development. *Mathematical Thinking and Learning*, *22*(4), 249–257.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77-80.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*, 431-467

Martinho, A., & Kacelnik, A. (2016). Ducklings imprint on the relational concept of "same or different". *Science*, *353*(6296), 286-288.

Marx, K. (1992). *Capital: Volume 1: A Critique of Political Economy* (Illustrated ed.). Penguin Classics.

Obozova, T., Smirnova, A., Zorina, Z., & Wasserman, E. (2015). Analogical reasoning in amazons. *Animal cognition*, *18*(6), 1363-1371.

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

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.

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*

Piaget, J. (1977). *The Essential Piaget* (J. J. Vonèche & H. E. Gruber, Eds.; 1st ed.). Basic Books.

Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences*, *6*(1), 125-136.

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). Sophia: 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*, 249–271.

Rogoff, B. (1981). Schooling and the development of cognitive skills. In H. C. Triandis & A. Heron (Eds.), *Handbook of cross-cultural psychology* (Vol. 4, pp. 233-294). Rockleigh, NJ: Allyn & Bacon.

Scribner, S. (1977). Modes of thinking and ways of speaking: Culture and logic reconsidered. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking* (pp. 483-500). Cambridge, England: Cambridge University Press.

Shakespeare, W. (1985). *Romeo and Juliet*. Ed. Durband, A., Hauppage, NY: Barron's.

Shakespeare, W. (1963). *As you like it*. Ed. Furness, H.H., New York: Dover Publications.

Shao, R., & Gentner, D., (2019). Symmetry: Low-level visual feature or abstract relation? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2790-2796). Montreal, QB: Cognitive Science Society.

Simms, N. K., & Richland, L. E. (2019). Generating relations elicits a relational mindset in children. *Cognitive Science*, *43*(10), e12795.

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows Spontaneously Exhibit Analogical Reasoning. *Current Biology*, *25*(2), 256–260.

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*

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.

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.

Vendetti, M.S., Wu, A. and Holyoak, K. J. (2014). Far-out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological Science, 25*, 1-6.

Vygotsky, L. S. (1929). The Problem of the Cultural Development of the Child. *Journal of Genetic Psychology, 36*, 415-434.

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological science, 25*(1), 161-169.

Walker, C. M., Bridgers, S., & Gopnik, A. (2016). The early emergence and puzzling decline of relational reasoning: Effects of knowledge and search on inferring abstract concepts. *Cognition*, *156*, 30-40.

Wasserman, E. A., Castro, L., & Fagot, J. (2017). Relational thinking in animals and humans: From percepts to concepts. *American Psychological Association Handbook of Comparative Cognition, Vol. 2,* J. Call (Editor-in-Chief), pp. 359-384.

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

Wittgenstein, L., (1977). *Culture and Value*. University of Chicago Press.

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.