Cognition 222 (2022) 105007

Contents lists available at ScienceDirect

Cognition

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

# You cannot find what you are not looking for: Population differences in relational reasoning are sometimes differences in inductive biases alone

Ivan G Kroupin *, Susan E Carey

*Harvard University, USA*

ARTICLE INFO

*Keywords:*
Relational reasoning
Analogy
Development
Human uniqueness

ABSTRACT

Relational reasoning is a cornerstone of human cognition. Extensive work, drawing on the Relational Match to Sample paradigm (RMTS), has established that humans, at least above the age of five, are much more proficient relational reasoners than younger children or non-human animals. While sometimes differences between populations derive from differences in capacity (the capacity to create representations in a certain format or of a certain complexity, information processing capacity), other times such differences derive from different learning histories alone. Here we distinguish between two types of learning history explanations on the example of four-year-olds' failure on Premack's (1983) RMTS task: (1) that children four-year-olds have not yet created representations of the relations same and different with the properties need to support success on RMTS and (2) that four-year-olds have different inductive biases than do adults. Experiment 1 established that four-year-olds are at chance on the RMTS task we deploy as a transfer task in Experiment 2. Experiments 2A–C each provide children with a mere 8 trials of training on of one three MTS tasks (Number, Size and Identity MTS, respectively), none of which involves making matches of same to same or different to different. The very brief training (eight trials) on two of these tasks (Number MTS, Size MTS) leads to spontaneous success on RMTS in four-year-olds. Identity MTS has no effect on subsequent performance on RMTS. Given the brevity and non-relational nature of the training the successes after Number and Size MTS training must have resulted from changing inductive biases alone. Furthermore, the same two training tasks increased relational responding by adults on a related task (Kroupin & Carey, in press), whereas Identity MTS training did not, suggesting that the mechanisms through which the training changed inductive biases are at least partially continuous between ages four and adulthood.

## 1. 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. Halford, Wilson, & Phillips, 2010; Holyoak & Thagard, 1995; 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 (1595/1985) 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?

### 1.1. 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 Fig. 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.

Even in the case of such an apparently simple instance of relational

* Corresponding author at: Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138, USA.
*E-mail address:* ikroupin@g.harvard.edu (I.G. Kroupin).

https://doi.org/10.1016/j.cognition.2021.105007
Received 18 February 2020; Received in revised form 14 December 2021; Accepted 23 December 2021
Available online 3 January 2022
0010-0277/

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, in press 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, Castro, & Fagot, 2017 for a review of the comparative literature).

### *1.2. Plan of the paper*

The current paper seeks to 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, in press). 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* 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 having the representational and computational capacities to succeed, *and* despite having already formed representations of the relations same and different sufficient for success. 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.

### *1.3. 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).

#### *1.3.1. Capacity limitation accounts of population differences in RMTS performance*

*1.3.1.1. Account 1 – limited representational capacity.* Numerous authors (e.g., Penn, Holyoak, & 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).

*1.3.1.2. Account 2 – 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 vary as a function of the nature of same/different representations used in completing the task (e.g. Thompson, Oden, & Boysen, 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).

#### *1.3.2. 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 et al., 1997). Wasserman et al. (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. Christie & Gentner, 2014; Kotovsky & Gentner, 1996). Even infants can successfully discriminate between same and different pairs (e.g., Walker & Gopnik, 2014; Ferry, Hespos & Gentner, 2015; Hochmann, Hochmann, Carey, & Mehler, 2018). Such evidence (discussed further below) suggests that children under the age of five have the *capacities* to succeed on standard RMTS (i.e. versions of the task where both same *and* different are included as bases of matching). However, to date, there have been no published data confirming this

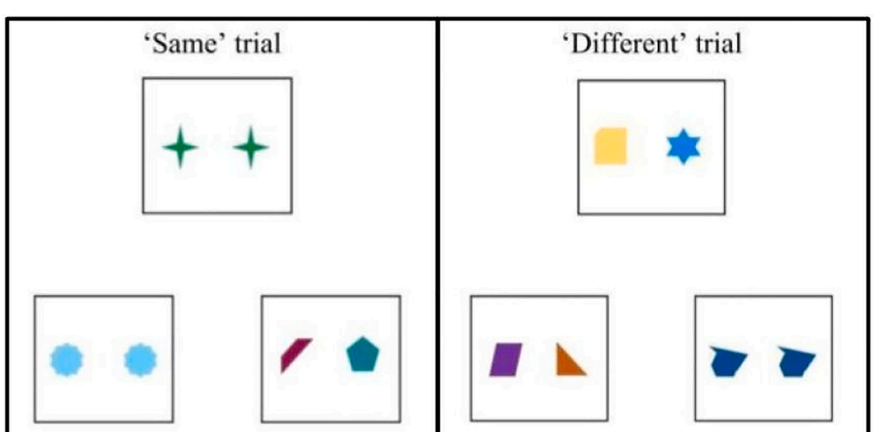


**Fig. 1.** Examples of two RMTS trials.

possibility. In sum, capacity limitations accounts of population differences in standard RMTS performance have been *ruled out* for some cases of failures of some 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 alternative accounts of persistent failures on RMTS, in addition to capacity limitations.

#### 1.3.3. Learning experience accounts of population differences in RMTS performance

*1.3.3.1. Account 3 – absence of specific representations.* One alternative account of failures at RMTS (e.g. Christie & Gentner, 2014; Gentner, 1988) 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 – and it is even possible that the words are initially mapped to representations which are not in such a format (see below for a discussion of possible formats).

*1.3.3.2. Account 4 – differences in inductive biases alone.* Several recent papers (Carstensen et al., 2019; Kroupin & Carey, 2021, in press, 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.

Importantly, it is clear to us that *Accounts 1–3 are correct descriptions of some population differences in relational reasoning performance*. For instance, Amoeba and earthworms probably lack, in principle, the representational (Account 1) and computational (Account 2) capacities for RMTS; and if so no training regime could induce success on any version of a RMTS task. Likewise, it is clear that some populations fail some tasks due to *not having yet developed* the necessary representations, despite being capable of doing so (Account 3). For instance, children up until a certain age are unlikely to have had a chance to construct a representation 'east of'. Even if they were *capable* of such a construction children who had not actually constructed a representation 'east of' would, in line with Account 3, fail an RMTS task involving matching pairs of cities on the basis of whether they are both east of Chicago or not both east of Chicago (e.g. Does [Seattle, LA] go with [Minneapolis, Denver] or [Boston, Atlanta]?)

It is also clear that differing inductive biases can contribute to an individual's failing any given relational reasoning task. Indeed, Gentner and her colleagues have demonstrated that the salience of object matches contributes to children's failures in relational reasoning paradigms. But as we detail below, and as Gentner and her colleagues point out, these demonstrations are consistent with both Accounts 3 and 4. Our goal here is to explore account 4; the stronger hypothesis that in some cases, not only does learning change inductive priors, but also that population differences on relational reasoning tasks may *sometimes* reflect differences in inductive biases *alone.* Adjudicating between Accounts 3 and 4 is important because qualitatively different learning mechanisms are likely to be involved in the processes that result in conceptual construction (Account 3: e.g., Quinian bootstrapping, Carey, 2009, conceptual modelling practices that involve structure mapping e.g. Gentner, Brem, Ferguson, & Wolff, 1997) and those that involve changing inductive biases alone (e.g., Bayesian updating, associative mechanisms; see Bascandziev, Tardiff, Zaitchik, & Carey, 2018, for evidence of qualitatively different learning mechanisms involving changes that fall under Accounts 3 and 4).

### 1.4. Discriminating between learning experience accounts (3 and 4)

#### 1.4.1. The ambiguity of relational training and labelling

The success of non-human animals and young children on RMTS as a result of some training paradigms rule out capacity accounts of population differences – 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 et al., 1997, the words "same"/"different" or a novel noun in Christie & Gentner, 2014). Both strategies can be explained by *either* learning experience account (3 or 4): 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, *alone,* so as to make matching according to relations *more likely to be inferred as relevant* in the context of the task (see Kroupin & Carey, 2021, in press for extended discussions).

#### 1.4.2. The Smirnova, Zorina, Obozova, and Wasserman (2015) paradigm

The most promising paradigm for distinguishing between the two learning experience accounts of population differences was introduced in studies with crows and parrots by Smirnova et al. (2015) and Obozova, Smirnova, Zorina, and Wasserman (2015), respectively. Specifically, training in this paradigm allowed birds to succeed, spontaneously (i.e., with no error feedback), on *three* separate RMTS tasks: Size RMTS, Color MTS and Shape RMTS (matching by same-size, same-color and same-shape, respectively). 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, Fig. 1), birds received three differentially-reinforced trials in which relational matches were *also* object matches (e.g. DD goes with DD or EF).

As Gentner and her colleagues point out, 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, consisted 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,[1] then number and finally size (see Smirnova, Obozova, Zorina, & Wasserman, 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 (Account 3). 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*.

#### 1.4.3. Evidence that MTS training changes inductive biases in adults

Previous work from our lab (Kroupin & Carey, in press) 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), 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.

### 1.5. 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, in press) will produce spontaneous RMTS success in a population which ordinarily fails the task – in this case four-year-old children. Importantly, the MTS training tasks do not involve progressive alignment, nor any other training that directly involves matching on the basis of 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.

#### 1.5.1. Causal RMTS/same-different discrimination

In a 'causal RMTS' paradigm, Walker and Gopnik (2014), also see Walker, Bridgers, & Gopnik, 2016) have demonstrated that children as young as 18-months 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 (a *same* pair, e.g. AA) are placed on top of the box it lights up, but not when two distinct objects are placed on top (a *different* pair, e.g., BC). For another group of children, it is the different pair that activates the machine. Then, when presented two new pairs of objects (e.g. EE, FG) and asked "to make the machine go", they chose the pair that instantiates the same relation within the demonstrated pair that did so. 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., 2018) 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, Mody, and Carey (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, Vijayan, Rao, and Vishton (1999) "rule abstraction" experiments (identifying the similarity among "la di la", zu mo zu, te pa te" and distinguishing such triads from "di di gu" while generalizing the pattern to "di gu di" (see Hochmann et al., 2018; Hochmann, under review, 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.

#### 1.5.2. 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. For instance, four-year-olds failed an RMTS task where matching by symmetry involved mapping across dimensions – (e.g. from size to shading, does aAa go with **OO**O or **O**O**O**), unless they had previously succeeded on mapping symmetry within dimensions – e.g. size to size, bBb to cCc or Ccc). While this is certainly evidence that children can match over the dimensions of symmetry and monotonic increase, Gentner and her colleagues (Kotovsky & Gentner, 1996, see also Gentner, Anggoro, & Klibanoff, 2011) have themselves repeatedly

[1] i.e. matches on all dimensions. In fact the Smirnova et al. (2015) paper describes them as 'shape' matches, but the objects on sample and correct choice cards in fact corresponded on all of shape, color and size.

acknowledged that the effect of progressive alignment can be due to two different mechanisms: One the one hand, it may have served to *develop* representations of these dimensions de novo in children (Account 3). On the other, it may have changed inductive biases so as to make children more likely to infer them as correct bases of matching. The latter possibility, if it occurs *without* representational change, corresponds to Account 4.

### *1.5.3. 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 "training children with relational labels for same and different makes the identity relation more salient, allowing them to perceive the commonality between two instances of this relation." (p. 391). Here they are appealing to a change in inductive biases. However, it remains a clear possibility (and indeed Christie and Gentner assume) that this experiment contributed to teaching children the words "same" and "different." Such learning involves producing representations of these relations in a format that is better suited to RMTS than are the representations of these relations available prior to learning the words. Consequently, these results cannot clearly distinguish between a learning experience account involving changes to inductive biases alone (Account 4) and one in which there are also changes in the nature of the representations of the relations same and different (Account 3). We agree with Gentner and her colleagues that both Accounts 3 and 4 are correct, for different episodes of learning, and sometimes within a single episode of learning. Our goal here is to explore whether one well-documented population difference may reflect differences in inductive biases alone (Account 4).

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 promotes 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.

To recap, the present studies test two hypotheses: 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 (in press), i.e. Number and Size MTS will, but Identity MTS will not. This would provide evidence that the mechanisms by which MTS training tasks have this effect are at least somewhat continuous across development.

## 2. 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 and Carey (in press). 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.

### *2.1. Participants*

Participants were 24 English-speaking children aged 49–60 months (M = 54.23 m, nine girls, fifteen boys) recruited by phone from the greater Boston/Cambridge area or at a local children's 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 sticker or small prize and a high-five for participating. Families who were tested in the lab were also given five dollars of travel compensation.

### *2.2. Materials*

Each RMTS trial contained three laminated paper cards, each of which contained two geometric figures (Fig. 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.

### *2.3. 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.

### *2.4. 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 RMTS tasks by children below the age of five (Hochmann et al., 2017; Premack, 1983). 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, $p = .03$, $t = 2.24$, Fig. 2). 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. Hochmann et al., 2017; Premack, 1983). 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.[2] 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.

## 3. 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 used with adults in Kroupin & Carey, in press). 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.

[2] 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 ($t \geq 1.7$, $p \geq .1$). 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, $t = 0.86$, $p = .39$). 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.

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, in press) 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.

### 3.1. *Experiment 2A: number MTS training*

#### 3.1.1. *Participants*

Participants were 24 English-speaking children aged 48–60 months (M = 53.57 m, 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.

#### 3.1.2. *Materials*

*3.1.2.1. Training task – number MTS.* Each Number MTS card contained one or three geometric figures (Fig. 3). 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.

*3.1.2.2. Test task – RMTS.* The RMTS stimuli were identical to those in Experiment 1.

#### 3.1.3. *Procedure*

*3.1.3.1. 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.

*3.1.3.2. 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.

#### 3.1.4. *Results*

*3.1.4.1. 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

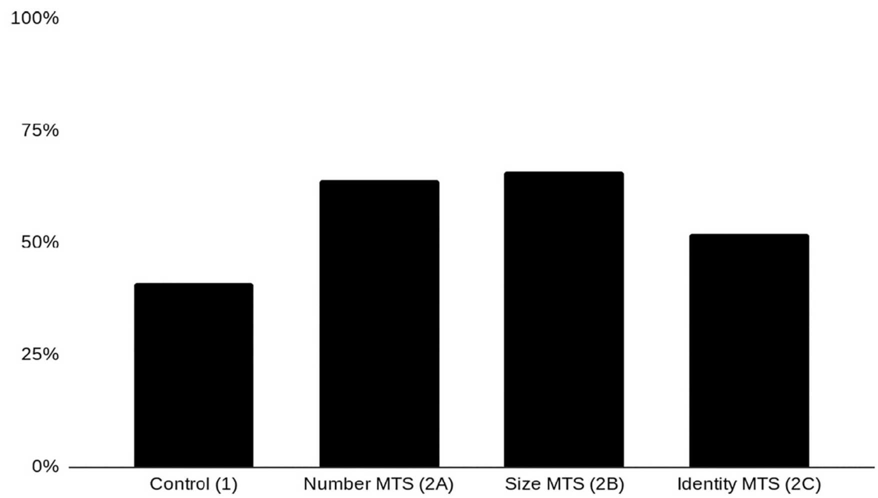


**Fig. 2.** Overall percentage correct on RMTS trials by children in each of Experiments 1–2C. The x-axis displays the training task used (if any) with the experiment number in parentheses.

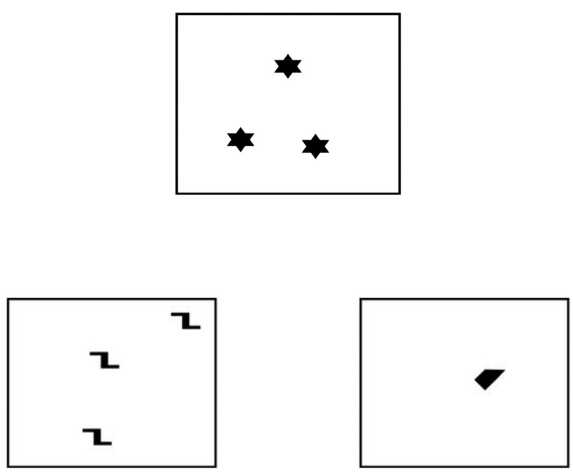

**Fig. 3.** Number MTS trial.

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, in press). 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. Chan & 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.

*3.1.4.2. 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, $p < .01$, $t = 2.91$). Of course, 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, $p < .001$, $t = 3.67$, Fig. 2. Furthermore, the improvement in percentage of relational matches was entirely due to the 7 spontaneous succeeders (i.e. children matching 7+/8 trials correctly on RMTS), who made relational matches on 94.6% of the trials. The children who did not succeed spontaneously made the relational match, on average, 51% of the time (i.e. were at chance, $p = .68$, $t = 0.42$).

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, in press). 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.

### *3.2. Experiment 2B: size MTS training*

#### *3.2.1. Participants*

Participants were 24 English-speaking children aged 49–60 months (M = 53.20 m, 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.

#### 3.2.2. Materials

*3.2.2.1. 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; Fig. 4). 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.

*3.2.2.2. Test task – RMTS.* The RMTS stimuli were identical to those of Experiment 1.

#### 3.2.3. 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.

#### 3.2.4. Results

*3.2.4.1. Training task – size MTS.* Only 4 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, in press). 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 real-world size is encoded automatically in the visual system, even by age four (e.g., 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.

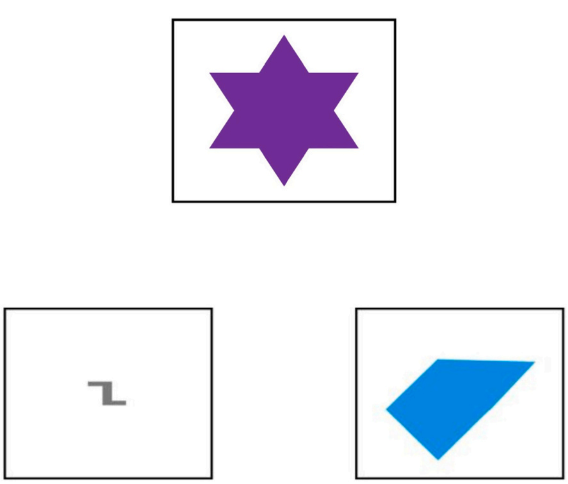

**Fig. 4.** Size MTS trial.

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.

*3.2.4.2. Test task – RMTS.* In spite of receiving no correction on the RMTS test trials whatsoever, not even error correction, 5 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, $p < .001$, $t = 3.60$, Fig. 2). Of course, as a group, children made also more relational matches than those in Experiment 1 (66% v. 41% relational matches, $p < .0001$, $t = 4.19$). Again, the improvement in percentage of relational matches (relative to chance or to baseline) was mainly due to the 5 succeeders (who made relational matches on 100% of trials). The non-succeeders were only marginally more likely to make relational matches than chance (57%, $p < .06$, $t = 1.97$).

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 and Carey (in press), 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.

### 3.3. Experiment 2C: identity MTS training – RMTS test

Unlike Number and Size MTS, Identity MTS did not increase relational responding in adults (Kroupin & Carey, in press). 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).

#### *3.3.1. Participants*

Participants were 24 English-speaking children aged 49–59 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.

#### *3.3.2. Materials*

*3.3.2.1. Training task – identity MTS.* Each Identity MTS card displayed one geometric figure (Fig. 5). 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.

*3.3.2.2. Test task – RMTS.* The RMTS cards were identical to those in Experiment 1.

#### *3.3.3. 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.

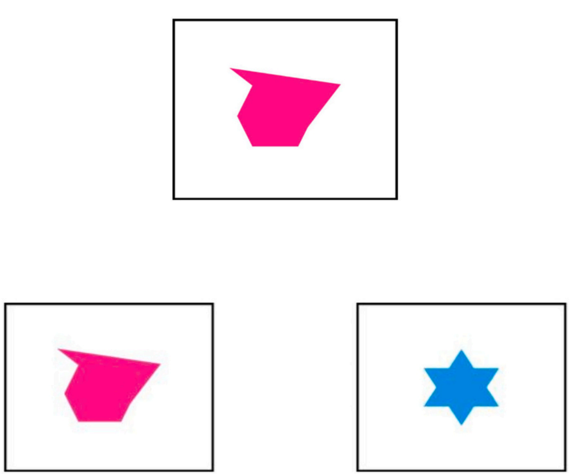

**Fig. 5.** Identity MTS trial.

#### *3.3.4. Results*

*3.3.4.1. 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/or color (or some measure of overall similarity which includes shape and/or color) to be correct bases of matching. 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, in press). 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, consistent with a large literature showing increased spontaneous attention to numerosity across age in Western samples (e.g. Chan & Mazzocco, 2017; McMullen, Verschaffel, & Hannula-Sormunen, 2020) as well as increased spontaneous attention to the relations same and different with age in on RMTS (Experiment 1, Hochmann et al., 2017; Premack, 1983).

*3.3.4.2. 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, $p = .03$, $t = 2.31$, Fig. 2), 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.

## 4. 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.[3] 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 of 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 four-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 at least in part one in inductive biases *alone*. However, as discussed above, there 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?

### *4.1. 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) or 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 (e.g. Diamond, Barnett, Thomas, & Munro, 2007). And by definition, any actual representation the child draws upon was within 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 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." Hochmann et al. (2016) and Zentall, Andrews, and Case (2018) provide evidence that it is exactly these procedures that underlie fourteen-month-old infants' and pigeons' MTS and nMTS performance. 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.[4] 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 $\Omega$). If children in Experiments 2A and 2B *already had* these representations, MTS training tasks did not *produce* them. If these same children *did not* have these representations, 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 these eight* instances of matching (i.e. Number or Size MTS in our lab) and not the vast number of instances of matching they are likely to have experienced in their lives up until participating in this study (e.g., every time they recognize something, categorize something).

#### *4.1.1. 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 (Smirnova 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 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

[3] Converging with this conclusion, Shao and Gentner report in unpublished data that four-year-olds spontaneously succeed on standard RMTS task (i.e. with both same and different sample cards). Given the repeated failures reported at this age in the literature (Premack, 1983; Hochmann et al., 2017; Experiment 1) it follows that this success must have been a result of some difference in the stimuli used by these authors which made children more likely to spontaneously infer sameness and difference as correct bases of matching. Further research should explore what this difference is.As such, this success, if confirmed, would only provide more evidence for our proposal that some population differences in relational reasoning are the result of differences in inductive biases alone. Moreover, work from our lab (Kroupin, under review) provides direct confirmation of the role of stimuli differences by showing that a targeted modification of the RMTS stimuli (by making them vary only on size, so as to inhibit partial matches on shape and/or color) leads to spontaneous success in four-year-olds.

[4] 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.

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.

### 4.2. *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.

#### 4.2.1. *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.

#### 4.2.2. *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, the 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 why some four-year-olds fail on RMTS despite Number/Size MTS training in more detail. Moreover, future work should investigate the relation of knowing the words "same" and "different" to RMTS success. After all, 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."

### 4.3. *Investigating inductive biases*

#### 4.3.1. *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 (in press). 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 and Carey (in press) 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 with respect to RMTS in adults.

#### 4.3.2. *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. all of 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. Chan & Mazzocco, 2017; Christie & Gentner, 2010).

These results are consistent with a large body of findings indicating that children initially prefer to make matches on object identity over relational matches (e.g. Gentner, 1988; Gentner & Namy, 1999; Gentner & Rattermann, 1991; Gentner & Toupin, 1986; Paik & Mix, 2006; Richland, Morrison, & Holyoak, 2006; Thibaut, French, & Vezneva, 2010). That being said, the present results clearly suggest that the object bias is not a general bias to weigh all features of objects more heavily than all relations between objects with respect to a given matching task, given the dramatic difference in rates of spontaneous success between Size and Identity MTS tasks in Experiment 2. We return to the issue of the domain-generality of inductive biases relevant to RMTS in more detail below.

Given there are no perfect matches on shape and/or color on RMTS (Fig. 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, in press). Fig. 6 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

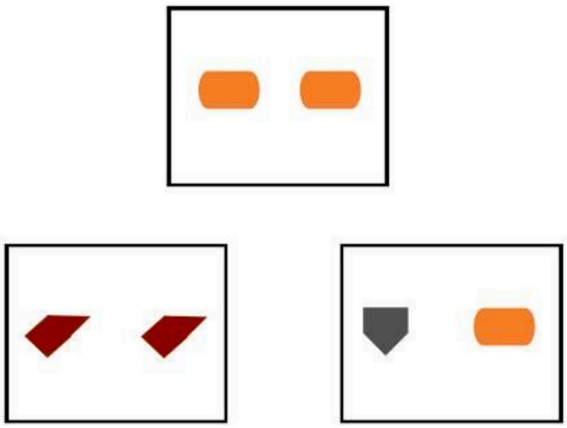

**Fig. 6.** A trial of the modified RMTS task used with adults in Kroupin and Carey (in press).

training tasks may have changed inductive biases so as to increase the salience of relational matches relative to shape/color matches.

#### *4.3.3. 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*; that is, some MTS training tasks (in this case Number or Size MTS) 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* – some MTS training tasks (in this case Number or Size MTS) 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.

#### *4.3.4. Number MTS*

*4.3.4.1. 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 Fig. 3) 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).

*4.3.4.2. 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 (unlike in 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.

### *4.4. Size MTS*

*4.4.1.1. 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, Kroupin (under review) uses a Size RMTS task in which neither shape nor color are available as bases of matching. That is same-different relations are conveyed in terms of the size of object (same-size v. different-size) but matches on color and shape are equated across choice cards (see Fig. 7). If four-year-olds' failure on standard RMTS is due to inductive biases which lead them to prefer shape and color *specifically*, children should be more likely to *spontaneously* infer the relations same and different as correct in such a task where shape and color are made less plausible as bases of matching – which is, in fact, the result Kroupin (under review) reports: Four-year-olds succeeded on a size-only RMTS task with no prior training *of any kind*. Such spontaneous success is consistent with

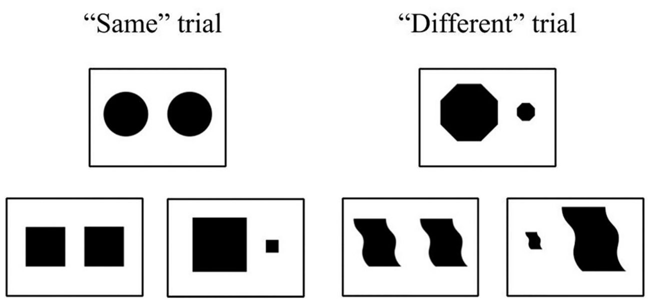


**Fig. 7.** Example RMTS trials from Kroupin (under review).

the hypothesis that training which decreases likelihood of matching on the basis of shape and color would increase likelihood of matching on the basis of the relation same in standard RMTS (Fig. 1).

*4.4.2. 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, in press) (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 (unsuccessfully) *attempt* partial shape/color matches should increase relational responding. This is, in fact, the case with adults (Kroupin & Carey, in press).

### 4.5. Match computations v. representations of the relations same and different

Unlike Number and Size MTS, Identity MTS training did *not* increase four-year-olds RMTS performance above chance. This result is important since it confirms the argument first made by Premack (1983) and elaborated by Hochmann et al. (2016) and Zentall et al. (2018) that the computations required to identify sameness in MTS *are not* equivalent to the kind of same/different representations required for RMTS success. After all, if match computations were the same as same/different representations, Identity MTS should have robustly increased four-year-olds' RMTS performance by promoting matches on the latter. Thus, evidence of success on MTS is *not* evidence of possessing the kind of same/different relations required for RMTS. A moment's reflection should confirm that evidence of successful *matching* is not evidence of the capacity to represent or reason with *relations*. Any creature capable of recognition or categorization carries out match computations. But not all such creatures can solve same/different discrimination tasks, let alone RMTS.

### 4.6. How general are inductive biases relevant to relational reasoning?

It is important to note that the mechanisms we are proposing do not presuppose that there are differences in inductive biases at a domain formulated at a domain-general level of description, such that some individuals selectively attend to object features and others to relations, across the board (as suggested by Gentner's notion of an 'object bias', e.g. Gentner, 1988; Gentner & Rattermann, 1991 as well as in more recent work, e.g. Carstensen et al., 2019, Simms & Richland, 2019; Vendetti, Wu, & Holyoak, 2014). Rather, we are assuming that *some* object features (shape/color) of some stimuli (geometric figures) are considered more relevant bases of matching than *some* relations (same/different), while other object features (size) are not. This proposal is supported by the results from the size-only RMTS task (Fig. 7) from Kroupin (under review) described above. That is, if children's initial failures on RMTS (as in Experiment 1) were a result of being more likely to make partial matches on *any* object features over *any* relational features, the modification of stimuli in this way should not affect performance – after all partial size matches are still available in the task as are, in principle, partial matches on shape or full matches on color. Yet, children show robust, spontaneous success on the modified RMTS task, supporting the idea that the pre-existing inductive biases guiding their relational reasoning are specified over *particular dimensions* of stimuli (at least in RMTS tasks, see also Kroupin, under review, for a detailed discussion of this issue). The results of Experiment 2 above and Kroupin (under review) indicating a differential weighting of color and shape (relative to other object features) also converge with findings on children's similarity judgements, namely that shape and color are weighted most heavily by four-year-olds in Western samples (Chan & Mazzocco, 2017).

### 4.7. 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 alongside* 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 alone can lead to a misidentification of inductive failures as limitations in the *representations* or *capacities* required to reason relationally. Second, as perhaps should be obvious, 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.

## Declaration of competing interest

The authors have no conflict of interest to declare.

## Acknowledgements

The authors would like to thank the many undergraduate research assistants who helped in collecting and coding the data presented here, first and foremost Sienna Nielsen, as well as financial support from Harvard University sources and the James S. McDonnell foundation.

## References


Bascandziev, I., Tardiff, N., Zaitchik, D., & Carey, S. (2018). The role of domain-general cognitive resources in children's construction of a vitalist theory of biology. *Cognitive Psychology, 104*, 1–28.

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

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

Carstensen, A. B., Zhang, J., Heyman, G. D., Fu, G., Lee, K., & Walker, C. M. (2019). Context shapes early diversity in abstract thought. *Proceedings of the National Academy of Sciences, 116*(28), 13891–13896.

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.

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.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387.

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.

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

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

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure-mapping and relational language support children's learning of relational categories. *Child Development, 82* (4), 1173–1188.

Gentner, D., Brem, S., Ferguson, R., & Wolff, P. (1997). Analogy and creativity in the works of Johannes Kepler. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 403–459). American Psychological Association.

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

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development, 14*(4), 487–513.

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

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10*, 277–300.

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.

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., Carey, S., & Mehler, J. (2018). Infants learn a rule predicated on the relation same but fail to simultaneously learn a rule predicated on the relation different. *Cognition, 177*, 49–57.

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. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

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

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. (in press). Inference in relational reasoning: Relational-match-to-sample as a case study. Journal Experimental Psychology: General.

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.

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

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

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

Paik, J. H., & Mix, K. S. (2006). Preschoolers' use of surface similarity in object comparisons: Taking context into account. *Journal of Experimental Child Psychology, 95*(3), 194–214.

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

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

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.

Shakespeare, W. (1985). In A. Durband (Ed.), *Romeo and Juliet*. Hauppage, NY: Barron's (Original work published 1595).

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

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology, 25*(2), 256–260.

Smirnova, A. A., Obozova, T. A., Zorina, Z. A., & Wasserman, E. A. (2021). How do crows and parrots come to spontaneously perceive relations-between-relations? *Current Opinion in Behavioral Sciences, 37*, 109–117.

Thibaut, J. P., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology, 106*(1), 1–19.

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.

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

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.

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

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

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.


## Data reference


???[dataset] 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, OSF.io, 2021, DOI 10.17605/OSF.IO/M5UAV.