



HHS Public Access

Author manuscript

Vision Res. Author manuscript; available in PMC 2018 December 01.

Published in final edited form as:

Vision Res. 2017 December ; 141: 228–236. doi:10.1016/j.visres.2016.12.005.

Hometown size affects the processing of naturalistic face variability

Benjamin Balas and

Department of Psychology, Center for Visual and Cognitive Neuroscience, North Dakota State University.

Alyson Saville

Abstract

Face exposure during development determines adults' abilities to recognize faces and the information they use to process them. Individual differences in the face categories represented in the visual environment can lead to category-specific deficits for recognizing faces that are atypical of observer's experience (e.g. the other-race effect). But what happens when observers have limited opportunities to learn about faces in general? In previous work, we found that observers from depopulated areas have poorer face recognition performance than observers from larger communities, suggesting that impoverished face experience limits face processing broadly. Here, we further investigate this phenomenon by examining how hometown size impacts the ability to assess appearance variability in natural images of faces and bodies. We asked individuals from small and large communities to complete (1) an unconstrained card-sorting task designed to test observers' ability to categorize within-person and between-person appearance variability properly, and (2) the Cambridge Face Memory Test. For both tasks, we examined the direct comparison between groups as well as the relationship between CFMT scores and sorting performance as a function of face experience. We find that small-town observers perform more poorly on the CFMT, but exhibit both better and worse performance than large-town observers on different aspects of the card-sorting task. Further, we also examine the relationship between CFMT performance and card-sorting errors. Our results suggest that individual differences in lifetime face exposure induce important variation in face processing abilities.

Keywords

Face recognition; visual development; unfamiliar faces; individual differences

Correspondence regarding this article should be addressed to Benjamin Balas, Department of Psychology, North Dakota State University, Fargo, ND 58102 Contact: Benjamin.balas@ndsu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Individual differences in observers' visual experience with faces lead to variation in their ability to recognize them. Perhaps the most profound example of this comes from multiple studies describing the various deficits in face processing that observers who were born with congenital cataracts experience even years after their cataracts have been removed. Typically, these participants have had their cataracts treated relatively early in infancy, yet in multiple tasks it is evident that their face recognition abilities differ from typical observers. For example, patients treated for bilateral congenital cataracts perform more poorly on the Famous Faces task, the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), and the Benton Facial Recognition Task (Benton et al., 1983; de Heering et al., 2011), despite self-reporting face recognition skills in the same range as typical observers (de Heering & Maurer, 2014). These assessments involve recognizing ostensibly familiar individuals (the Famous Faces Task), learning to recognize novel faces (the CFMT) and matching identity across changes in view using simultaneously presented images (the Benton Facial Recognition Task). Cataract-reversal patients thus exhibit a number of important deficits that suggest that their disrupted early experience with faces limits their ability to effectively recognize and discriminate individual faces. Besides this evidence that cataract-reversal patients tend to perform less accurately across a range of face recognition tasks, there is also substantial evidence that the way they process faces is also different. Cataract-reversal patients appear to be worse than control participants at discriminating faces that differ according to changes in face geometry (e.g. eye spacing), but perform comparably when faces differ according to the local shape of intuitive facial features like the eyes, nose, and mouth (Le Grand et al., 2001; Mondloch, Robbins & Maurer, 2010). This suggests that the computation of the visual features (whatever they may be) that support the discrimination of faces based on these 2nd-order configural properties (Maurer, Le Grand & Mondloch, 2002) may be specifically impaired in these patients due to their impoverished early experience. Similarly, cataract-reversal patients also exhibit a reduced Composite Face Effect (Young, Hellawell, & Hay, 1987), which has been widely used as a proxy for "holistic" face processing. In the typical CFE, observers are asked to match or discriminate only the top half of face patterns while disregarding the bottom half of the stimulus. Observers tend to find this difficult, ostensibly due to obligatory processing of the entire face pattern which leads to interference from the bottom half of the stimulus even though it is task-irrelevant. If the bottom half of the stimulus is misaligned with the top half, the interference effect is reduced or absent. Unlike control participants, however, cataract-reversal patients do not appear to suffer from such interference (Le Grand et al., 2004), which suggests that they may not engage in obligatory holistic processing to the same degree as typical observers. Varying experience with faces thus influences the manner in which faces are processed as well as participants' ability to recognize them effectively. Critically, it is not the case that these patients have a broad visual recognition deficit. Their performance in a number of closely related tasks, including human face detection (Mondloch, Le Grand & Maurer, 2003) and recognition tasks using other complex patterns (e.g. monkey faces and houses - Robbins et al., 2010), is comparable to controls. This pattern of results indicates a specific relationship between individual differences in face experience and varying face recognition abilities.

True visual deprivation is a fairly extreme example of individual variation in visual experience with faces, and most people do not ever experience such deprivation during their lifetime. Moreover, though cataract-reversal patients do exhibit specific deficits in human face processing, their visual experience has been affected broadly; Pattern vision is broadly compromised prior to treatment. Presently, we chose to examine a subtler form of individual variation in face experience that does not depend on visual deprivation *per se*, but instead is a function of the visual environment: How does face recognition differ as a function of the number of faces you are exposed to? Specifically, does growing up in a sparsely populated area lead to poorer face recognition abilities relative to someone who grew up in a densely populated community? In a previous report (Balas & Saville, 2015), we demonstrated that adult observers who grew up in small communities (populations <1000) were less accurate than participants from larger communities (populations >30,000) at learning to recognize new faces, and had weaker face selectivity at the N170 component, which is a known marker of face processing in ERP signals (Rossion & Jacques, 2008). In this study, we characterized participants' ability to learn new faces using the Cambridge Face Memory Test (or CFMT, Duchaine & Nakayama, 2006), which is a standardized face memory assessment that has been widely used to characterize performance in individuals with prosopagnosia. Using the CFMT, we found that the scores of individuals from small-towns were significantly lower (~10%) than the scores of individuals from large-towns. In terms of the neural responses measured in both groups at the N170 component, we found that while large-town observers had robust differences between face and non-face amplitudes (which is typical of N170 response properties), small-town observers exhibited a smaller amplitude difference between faces and chairs, which suggests poorer category selectivity at this particular component. These results suggest that like biased experience favoring own-race faces over other-race faces (de Heering et al. 2010), the overall amount of face exposure observers receive during development affects their ability to recognize faces of all categories effectively.

Presently, we extend this line of work to examine how observers from “small-town” and “large-town” communities process naturalistic variability in images of faces and bodies. Effective person recognition depends critically on being able to tell the difference between image variation that occurs without a change in identity (intra-personal variation) and image variation that does result from a change in identity (extra-personal variation), and explicit modeling of these two sources of variability is the basis of successful computer vision systems for face recognition (Moghaddam, Jebara, & Pentland, 2000) and accounts for some features of infant face learning (Balas, 2012; Balas, 2013). Natural images of faces and bodies are highly variable, and while observers are generally able to cope with this high variability when asked to recognize familiar individuals (Bruce et al., 2001), they are generally quite poor when asked to match or discriminate unfamiliar individuals (Bruce et al., 1999; Bruce et al., 2001; Johnson & Edmonds, 2009). In particular, a series of results using a simple unconstrained card-sorting task (Jenkins et al., 2011), reveals key aspects of how observers fail to process intra- and extra-personal variation in images of unfamiliar people. Briefly, when asked to sort a set of images containing multiple instances of an unknown number of individuals based on identity (e.g., estimate the number of unique people in the set), observers tend to substantially overestimate how many different people are being depicted. The pattern of sorting errors suggests that participants are especially

poor at “telling faces together” (Andrews et al., 2015), by which we refer to the ability to determine that different images of the same person belong in the same identity group. This tendency is further exacerbated when other-race faces are used (Laurence, Zhou & Mondloch, 2015), suggesting that one consequence of reduced experience with a set of faces is an increased tendency to “split” identities up when appearance varies. Based on these results, we hypothesized that observers raised in small communities might also have more difficulty “telling faces together” than observers raised in larger communities, as evidenced by increased errors when attempting to categorize intra-personal variation correctly. By including body stimuli in this study, we are also able to comment on several additional issues related to person perception. First, including bodies allows us to examine the generality of card-sorting behavior to a non-face object. Independent of community size, does the sorting bodies exhibit the same patterns of behavior as the sorting of faces? Second, if we are able to observe measurable differences in sorting behavior as a function of community size for faces, do those effects extend to bodies as well? If not, this would suggest that face recognition is uniquely impacted by the variation in experience that is a consequence of hometown size. Alternatively, experience with person recognition (combining face and body expertise) may lead to more general effects of community size on sorting. Finally, there is as yet very little information about how appearance variability is processed in bodies using unconstrained tasks like this, so our inclusion of this condition also provides novel data regarding body processing in naturalistic images.

We recruited adult observers (all college undergraduates) who hailed from both small hometowns and large hometowns to take part in the aforementioned card-sorting task using images of faces and images of bodies. We also asked these participants to complete the Cambridge Face Memory Test, to replicate and extend our prior results regarding observers’ ability to learn to recognize new faces as a function of varying face experience during development. We hypothesized that observers with relatively impoverished face experience would perform more poorly on the CFMT, and would be less likely to sort different images depicting the same unfamiliar face into the same identity group during unconstrained sorting. We further conjectured that this effect might not be evident for bodies, if indeed visual experience leads to a specific deficit for face recognition. Finally, we chose to examine the relationship between sorting performance and CFMT performance by investigating the correlations between sorting errors and recognition outcomes in these two tasks across both groups.

Methods

Participants

Our final sample was comprised of 18 participants (7 female) who grew up in communities with more than 30,000 individuals and 21 participants (7 female) who grew up in communities with fewer than 1,000 individuals. All participants were Caucasian individuals between the ages of 18-23 years of age who self-reported normal or corrected-to-normal vision. We recruited these individuals using the NDSU Undergraduate Psychology Study Pool, and all participants received either course credit or monetary compensation for taking part in the study. We obtained written informed consent from all individuals before the

experimental session began, and the NDSU IRB approved the research carried out as part of this study. We also recruited an additional 3 individuals in the large-town sample and one additional individual in the small-town sample who were excluded either for failure to complete the tasks (N=2) or because they self-identified as non-Caucasian. Both groups were largely comprised of first-year students (Small-Town: N=12; Large-Town: N=10), but both groups also contained several upper-year students such that the number of years spent at university varied within each group. Members of our small-town group thus had variable exposure to the larger population density offered by the Fargo-Moorhead metro area.

Stimuli

All participants took part in two sets of tasks: the Cambridge Face Memory Test (CFMT) and an unconstrained card-sorting task using images of unfamiliar faces and bodies. We administered the CFMT using the online testable.org interface (<http://www.testable.org>) and the images used in this task are described in Duchaine & Nakayama (2006).

The stimuli used in our face/body card sorting task were comprised of 20 unique images of four different unfamiliar women: Doutzen Kroes, Sylvie Van Der Vaart, Elsa Hosk, and Lara Stone – all of whom are Dutch models who were unfamiliar to our participants. The original images were obtained from Google Image searches using the models' names as keywords and we only selected images depicting the face and body of the target individual in a naturalistic setting (no staged photographs). We created face-only images by cropping the central portion of the face (retaining the natural jaw-line, but excluding the hair) and re-sizing the face to fit within a 3" × 3" card. We created Body-only images by deleting the cropped face portion from the original photograph and resizing the remaining portion to fit within a 3" × 3" card. We printed all of these images out in full-color and cut them to size for use in the sorting task (Figure 1). The four models we selected were all highly attractive and had similar hair color and body type, but within these constraints we attempted to choose a set of individuals that were not so similar as to preclude discrimination from one another. We also chose images of these individuals such that paraphernalia (sunglasses, clothing, jewelry, etc.) could not be used to match distinct images of the same person.

Our study differs from prior reports (Jenkins et al., 2011; Andrews et al., 2015) in that we chose to use four distinct identities rather than two in our stimulus set. We chose to do so largely because we were concerned that by choosing just two unfamiliar models we might unduly constrain performance as a function of the similarity between the two target individuals. For example, if these two individuals were extremely similar to one another, we might make it very difficult for participants to appropriately distinguish them at all. On the other hand, if they were too distinct from one another, "intrusion errors" in which different individuals were placed in the same pile may be artificially low. Further, using only two individuals makes it possible to sort cards based either on successfully accepting the similarity between an individual image and the members of an existing group or successfully rejecting an image based on its dissimilarity to the members of an existing group. Our larger group of target identities reduces the impact of these issues.

Procedure

Participants completed both the CFMT and the unconstrained card-sorting tasks within a 45-60 minute testing session. Task order was alternated so that within each participant group, approximately half of our participants completed the CFMT first while the remainder completed the card-sorting task first. Further, the order of face and body card sorting was also alternated across individuals so that faces were sorted first by approximately half of the participants in each group.

CFMT—The Cambridge Face Memory Test was administered on a 1200×800 MacBook Pro. Participants were seated a comfortable viewing distance from the display. We did not constrain head position, nor did we monitor eye movements. The CFMT is comprised of 72 items, divided across three distinct phases: Introduction/Identical Image, Novel Images, Novel Images with Noise. During the introductory phase, participants are shown images of 6 target individuals, and attempt to choose the image depicting a target individual from an array comprised of one target (using an image identical to one of the study images) and two distractors. In the “Novel Images” phase, participants must attempt to identify the image depicting one of the 6 target individuals from a 3-item array, but now novel images of the target individuals are used. Finally, during the “Novel Images with Noise” phase, the task is made more complex via the addition of white noise to the images depicted in the test array. Further details regarding the CFMT can be found in Duchaine & Nakayama (2006). Participants completed the task in a sound-attenuated testing room and were free to complete the multiple phases of the CFMT at their own pace.

Unconstrained Card-Sorting—We asked participants to complete an identity-sorting protocol using both our face-only and body-only images of unfamiliar individuals. Following the procedure described in Jenkins et al. (2011), we presented individuals with the full deck of cards or the current task (either faces or bodies) and asked them to sort these cards into piles based on the identity of the individuals depicted in the photographs. Participants were told that each pile should contain all of the cards that they believed depicted the same individual and that they could choose to make as many of these piles as they thought were necessary. We told participants to refrain from using non-diagnostic contextual information to make their decisions (e.g. similar locations) and also advised them that paraphernalia (clothing, glasses, etc.) would not be diagnostic of identity and that therefore they should not attempt to use these cues to perform the task. We also emphasized that there could be as many as 80 unique individuals in the set or just one individual. Participants were free to re-assign cards to piles throughout the task and could view as many items at a time as they wished.

Results

CFMT performance

We began by examining whether or not the current sample of small-town and large-town observers differed from one another in terms of their ability to learn new face identities during completion of the CFMT. We have previously reported that small-town participants

were significantly worse at the task than observers from larger communities (Balas & Saville, 2015), so we anticipated that performance would also be lower here.

First, we compared overall CFMT accuracy in both groups. This score represents participants' performance across all three phases of the study, and is the measure we used to characterize face recognition abilities in our previous report. An independent-samples t-test revealed significantly lower performance ($t(37)=-2.09$, $p=0.043$, two-tailed t-test) in the small-town group ($M=0.69$, $s.e.m.=0.027$) relative to the large-town group ($M=0.76$, $s.e.m.=0.024$).

We continued by examining performance in the separate phases of the CFMT, but excluded data from the introductory phase due to the potential for ceiling effects to complicate the interpretation of our results. The modal value for each group in the introductory phase was 100% (16 perfect scores in the large-town group, and 15 in the small-town group) and no participant scored worse than 82% in either group. We chose therefore to only conduct a more detailed analysis of the data from the final two phases of the CFMT. For each observer, we recorded the proportion correct in both the "Novel Images" phase and the "Novel Images with Noise" phase and analyzed these values using a 2×2 mixed-design ANOVA with task phase as a within-subjects factor and participant group as a between-subjects factor. This analysis revealed a significant main effect of task phase ($F(1,37)=8.99$, $p=0.005$, partial eta-squared=0.20) and a significant main effect of group ($F(1,37)=4.09$, $p=0.05$, partial eta-squared=0.10). The interaction between task phase and group was not significant ($F(1,37)=0.026$, $p=0.87$). We display the average performance across both test phases for small-town and large-town observers in Figure 2.

The main effect of task phase was driven by poorer performance in the "Novel Images with Noise" phase ($M=0.60$, $s.e.m.=0.027$) than in the "Novel Images" phase ($M=0.68$, $s.e.m.=0.026$). The main effect of participant group was the result of poorer performance by small-town observers ($M=0.59$, $s.e.m.=0.032$) compared to large-town observers ($M=0.69$, $s.e.m.=0.034$).

Card-Sorting Performance

The data from the card-sorting task yields two interesting dependent variables to consider: The number of groups participants make (Figure 3), and the sorting errors that are evident in the assignment of cards to groups. We analyzed the former using a linear mixed-models analysis with stimulus category (face/body), and participant group as fixed effects, and participant ID as a random effect (due to the expectation that participant variance is not hypothesized to contribute a predictable amount of variation to the data). We estimated the statistical significance of these factors, and the interaction of category and participant group, by carrying out likelihood ratio tests that compared the full model (with all factors included) to a model with the factor under consideration omitted.

We also conducted a preliminary analysis to determine whether or not task order (faces-first vs. bodies-first task order) impacted performance. This analysis revealed that there was no main effect of task order ($\chi^2(1)=0.229$, $p=0.63$) and also that task order did not interact with participant group ($\chi^2(1)=2.27$, $p=0.132$) nor did it interact with stimulus category

($\chi^2(1)=0.25$, $p=0.62$). Given that task order thus does not seem to affect group numerosity per this analysis, we continue by discussing the results of the simpler analysis in which we have collapsed over task order.

This analysis revealed significant main effects of stimulus category ($\chi^2(1)=4.305$, $p=0.038$) and participant group ($\chi^2(1)=3.91$, $p=0.048$). The former effect was the result of participants making more face groups than body groups, while the latter was the result of small-town participants making more groups than their large-town peers (Table 1). The interaction between these two factors was not significant ($\chi^2(1)=0.34$, $p=0.56$).

By itself, group numerosity has a number of important weaknesses as a measure of accuracy in this sorting task. Our results comparing group numerosity across participant groups suggest that small-town observers are worse at this task than large-town observers because they are further from the correct answer of four groups. However, we suggest that the composition of the groups is of critical importance to consider to contextualize this result. To consider a simple example, an observer who makes four groups that each contain images of all four individuals has gotten the target number of piles exactly right, but has also clearly made a number of sorting errors. By comparison, an individual who has made twice as many groups that are homogeneous with respect to identity has done more poorly with regard to the target number of groups, but intuitively has also exhibited some ability to distinguish between identities effectively. Thus, we continue by conducting a novel type of sorting error analysis to determine how observers from both types of community made errors related to incorrectly grouping individuals together and incorrectly separating images of the same person. We emphasize that we do not intend this to supplant the use of group numerosity to consider performance in this paradigm, but rather introduce this analysis to offer a complementary set of descriptors that allow us to quantify performance with regard to both kinds of errors people can make in this task.

To examine the composition of the groups formed by our participants, we quantified sorting errors using a combinatorial calculation that reveals the rate of both “Same-Person, Different-Group” errors and “Different-Person, Same-Group” errors. The first type of error represents a failure to “tell people together” (Andrews et al., 2015), as evidenced by images of the same individual being assigned to different groups. The second type of error represents failure to “tell people apart,” as evidenced by cards depicting different individuals being assigned to the same group. We calculated the rates of these errors by considering all possible *pairs* of cards in the full set, because each pair of cards represents a unique opportunity for a participant to make a mistake. If a pair of cards depicts different people, they contribute to the “Different-Person, Same-Group” error rate if they are assigned to the same pile. Similarly, if a pair of cards depicts the same person, they contribute to the “Same-Person, Different-Group” error rate if they are assigned to different piles. We converted the sorting solution arrived at by each participant in the face and body tasks to estimates of these two error rates by counting the number of incorrectly sorted pairs and dividing each error rate by the appropriate denominator given the total number of same-person and different-person pairs in the full set. We obtained the total number of same-person pairs by multiplying the number of distinct identities (4) by the number of ways to choose 2 items from 20 (“20 choose 2,” which is equal to 190) yielding a value of 760. We obtained the

total number of different-person pairs by computing the number of ways to choose 2 distinct identities from a set of 4 (“4 choose 2,” which is equal to 6) and multiplying this by the total number of ways to pair the 20 identities belonging to one individual with the 20 identities belonging to another (which is 400), yielding a value of 2400.

An important issue to consider with regard to our counting procedure is the extent to which we can make a strong statement regarding the independence of these two kinds of error rate. We argue that because the card-sorting task is unconstrained (and participants can re-arrange cards as they wish) these two error rates are theoretically independent from one another. That is, participants are always free to re-arrange cards to reduce error rates of both types, and making one kind of error does not force observers to make another type of error. For example any individual card can always be placed in a group by itself, or placed in a group with a different card depicting another individual. In practice, and given an existing partial card-sorting solution, the sorting of an individual card does potentially contribute to both error rates. We think this does not substantially affect the conclusions we draw from these analysis, but it is an important consideration for this method (or any other method) that is designed to quantify performance in this rich behavioral task. Critically, any effects of community size would not rely on the strength of error independence, since any combinatorial constraints on how these errors are computed apply to both groups.

We submitted each set of values to a 2×2 mixed-design ANOVA with stimulus category (face vs. body) as a within-subject factor and participant group as a between-subjects factor. We chose to analyze the two types of error separately because they are not fully independent, and therefore an omnibus ANOVA did not seem appropriate. In each case, we conducted a preliminary analysis to determine if task order (faces-first vs. bodies-first task order) affected error rates. For both kinds of error, we found that there was no significant main effect of task order (for “Same-Person/Different-Group” errors, $p=0.70$; for “Different–Person/Same-Group” errors, $p=0.38$). In both cases, we also found that task order did not interact with either participant group or stimulus category. As a result, to simplify our analysis of both error types, we present the results of a simpler ANOVA in which we have collapsed across task order.

Our analysis of “Same-Person/Different-Group” errors revealed neither a main effect of category ($F(1,37)=0.217$, $p=0.64$) nor a main effect of participant group ($F(1,37)=1.23$, $p=0.28$). The interaction between these two factors was also not significant ($F(1,37)=1.30$, $p=0.26$). Consistent with prior reports describing card-sorting performance for face images, the mean error across all cells was high ($M=0.74$, 95% CI=[0.70 0.78]), which indicates that observers have substantial difficulty “telling faces together.”

Our analysis of “Different-Person/Same-Group” errors revealed a far lower error rate ($M=0.11$, 95% CI=[0.09 0.14]), which is also consistent with prior reports that indicate observers make so-called intrusion errors at a low rate. More importantly, we also observed significant main effects of stimulus category ($F(1,37)=27.2$, $p<0.001$, partial eta-squared=0.42) and participant group ($F(1,37)=4.37$, $p=0.043$, partial eta-squared=0.11). The former result was driven by significantly higher error rates for bodies ($M=0.16$, s.e.m.=0.019) than for faces ($M=0.067$, s.e.m.=0.008). The main effect of group was driven

by a significantly higher error rate among large-town observers ($M=0.14$, $s.e.m=0.018$) than among small-town observers ($M=0.087$, $s.e.m.=0.016$). The interaction between these two factors was not significant, though we note that follow-up t-tests indicate that the group difference does reach significance in the Face condition ($t(37)=2.52$, $p=0.016$, two-tailed t-test) but not the Body condition ($t(37)=1.58$, $p=0.12$, two-tailed t-test) despite the fact that the former difference appears smaller in Figure 4. In the absence of a significant interaction, we do not wish to draw firm conclusions about this “difference of significances” but the higher variability in the body condition may be the reason this comparison is not significant when considered alone.

Correlations between CFMT and Card-Sorting Scores

Finally, to examine the relationship between card-sorting performance in both tasks and CFMT scores, we examined correlations between CFMT accuracy and all four sorting error rates across the entire sample of participants. Specifically, we combined our small-town and large-town participants into one aggregate sample ($N=39$) and determined the strength of the correlation between CFMT performance and both kinds of face/body error rates. This analysis is intended as an exploratory investigation of how the ability to cope with appearance variability in face and body images may relate to the ability to learn new faces using a participant sample in which we know there is substantial individual variability in performance that is related to hometown size. There are several reasons why this analysis should be taken as a preliminary step towards characterizing the relationship between these aspects of person recognition, including (but not limited to) the lack of data regarding the reliability of card-sorting performance. As such, we intend these results to serve as an interesting initial look at how recognizing variability in naturalistic images may relate to face memory and learning.

In Table 2, we list the correlation coefficients between CFMT accuracy and each error rate across the entire sample. We found that there was a robust correlation between CFMT accuracy and Same-Person/Different-Group error rates for face card-sorting ($p<0.001$), but that the other correlations would not survive correction for multiple comparisons. As a result, we conclude that face learning and memory may be most strongly related to sorting errors that reflect failures to “tell people together.” We note that this relationship is a negative correlation, which is consistent with the reasonable expectation that lower sorting error rates should predict higher CFMT accuracy. While we emphasize that these correlations should be taken as a preliminary look at the relationship between sorting performance and face learning, these results suggest that telling faces together may be a particularly vital constraint on overall fluency with face images.

Discussion

Our results demonstrate several interesting ways that person recognition differs as a function of lifetime face exposure. First, we replicated our previous results regarding small-town observers’ poorer performance in the CFMT relative to observers who grew up in more densely populated areas. Our comparison between the “Novel Images” and “Novel Images + Noise” condition also suggests that it is not the case that relatively impoverished experience

with faces makes small-town observers disproportionately sensitive to additive noise in face images. Rather, performance appears to be uniformly lower across phases of the CFMT, with the exception of the Introductory phase, in which both groups performed near ceiling. We emphasize that these results indicate that small-town observers are not prosopagnosic – their performance in the more difficult phases of the CFMT is well above that realized by prosopagnosic observers, and they evince no difficulties during the introductory period, whereas prosopagnosic observers frequently make errors during this phase (Duchaine & Nakayama, 2006). Nonetheless, our current CFMT data reinforces our prior claim that limited face experience during development does result in measurable face recognition deficits during adulthood.

The results from our unconstrained card-sorting task offer further, and more complicated, insights into the nature of visual recognition in small-town observers. Like results examining other-race identity sorting (Laurence, Zhou & Mondloch, 2015), our small-town observers did create more groups than our large-town participants, which could be interpreted as evidence of a poorer ability to “tell people together” (Andrews et al., 2015). That is, small-town observers may create more groups because they tend to assign different images of the same person to different identity groups at a higher rate than observers from large towns. However, we argue that group numerosity alone cannot be used to unequivocally decide if an observer's (or a group's) performance is more or less accurate because the composition of those groups matters a great deal. We thus presented a novel analysis of error rates that allowed us to provide a richer picture of how observers in our two groups sorted identities. This analysis revealed differences in the “Different-Person/Same-Group” error rates which indicated that small-town observers were actually *less* likely to make so-called “misidentification errors” or “intrusion errors” than large-town observers, and did not make more “Same-Person/Different-Group” errors than large-town participants. This differs from results obtained with other-race identity sorting, where no robust differences in the misidentification rate were observed (Laurence, Zhou & Mondloch, 2015). Moreover, it is intriguing that this is an instance of small-town observers performing *better* than their large-town counterparts: is limited experience with faces a benefit in some ways? While we have assumed here and in previous work that small-town experience represents a state of mild deprivation relative to large-town face exposure, historical (Dunbar, 1992) and modern (Hamilton et al., 2007) trends in the size of hunter-gather groups indicate that the range of population sizes we have considered as a “small-town” band (<1000 people) are actually very typical of hunter-gatherer societies and thus may be more representative of the conditions under which face recognition abilities were selected. Indeed, the mean size of personal networks in modern societies is also well within this range (McCarty et al., 2000) which may mean that even considering densely populated areas, the number of *meaningful* faces in the environment may be more consistent across observers than we have assumed, and this number may be a good bit smaller than the total population suggests. Individual differences in face recognition abilities may be more closely shaped by the number of people an observer actually interacts with rather than the total available population of faces in the environment, which we suggest is consistent with some developmental results. In infancy, exposure to other-race and other-species faces is not sufficient to extend infants’ abilities to discriminate between other-species faces beyond the period “perceptual narrowing,” but

exposure to the same faces with names attached does support superior performance (Scott & Monesson, 2009). Face learning in this context (and the maintenance of a representation that supports reliable face discrimination) depends on individuation rather than just the presence of more exemplars in the visual environment.

In light of this, it is interesting to consider our observation that CFMT performance is predicted by the rate of “Same-Person/Different-Group” face-sorting errors across our entire participant sample. Though this is admittedly very speculative, it does suggest that achieving face constancy (stable representations of individual identity across different images of the same person) is a key factor that constrains face processing. To put it another way, recognizing different identities may not be as fundamental to face recognition as being able to generalize correctly across varying images of the same person. While our current sample was not large enough to permit meaningful comparison of these relationships across our small-town and large-town groups, further work to determine how community or network size may impact these relationships would be another valuable way to characterize how variation in face experience modulates the representations and mechanisms that support person recognition. Also, in this instance (and for our other analyses as well) determining the descriptors of community or network size that are most relevant for investigating individual differences in experience on faces will likely be an important contribution. In particular, both diary studies of face experience (Rennels & Davis, 2008) and first-person video recordings of the visual environment (Sugden, Mohamed-Ali & Moulson, 2014) would be excellent means of obtaining better and more meaningful estimates of face experience in a diverse population of observers.

Regardless of these intriguing questions concerning the relevant group size we should consider when examining individual differences in faces experience, another important issue that our study does not allow us to address satisfactorily is what aspects of face experience actually drive poorer CFMT performance and differences between groups in unconstrained sorting. Specifically, we are unable to tell if the key difference between our participant groups is the sheer number of faces they are exposed to (or interact with), or differing amounts of face homogeneity in small-town vs. large-town environments. Within a racial category (e.g. Caucasian), face ethnicity can have an effect on face recognition performance such that own-race, different-ethnicity faces are more poorly recognized than faces that conform to the dominant regional ethnicity (Bowles et al., 2009). In particular, the CFMT is comprised of faces that are representative of Boston-area facial appearance, which makes it a useful instrument for many North American environments but less so for other regions. McKone et al. (2011) have shown that a complementary instrument (the CFMT-Aus), which is comprised of faces that are representative of Australian facial appearance, is a more useful diagnostic tool than the original CFMT for investigating Australian patients with developmental prosopagnosia. The communities that our small-town observers have lived in are likely to be very homogenous in terms of ethnicity, and German and Scandinavian ethnicity will be highly over-represented. One possible explanation for the current results, therefore, is that the real source of individual variation between small-town and large-town observers is the diversity of ethnicities participants were exposed to in their visual environment rather than the total amount of exposure. This is unfortunately rather difficult to test adequately; we would either like to find a community that is small, but very diverse, or

develop an instrument like the CFMT that is tailored to the specific ethnicities in rural North Dakotan communities. For the present, we will concede that while we have observed an interesting consequence of small vs. large-town face exposure, the dependence of these effects on face variability vs. population size is an open question.

One way to try and provide a coherent theoretical account of our results is to consider how community size might affect representations of identity within a face-space framework (Valentine, 1991). Briefly, a face-space is a dimensional model of facial appearance that makes it possible to consider individual instances of a face as points in some (usually high-dimensional) space. The set of images belonging to a single person is typically hypothesized to occupy some region of this space (a cell in a Voronoi tessellation of the space, e.g.) and various recognition tasks can in theory be accomplished by using geometric relationships between points and regions to make inferences about identity and category membership. In a space that is dominated by faces of one racial group, other-race faces tend to be outliers that occupy a cluttered region away from the densest part of the face-space representation. Laurence et al. (2015) have argued that the pattern of results observed for other-race card-sorting is consistent with this due to the way identity-specific regions would be defined under these conditions. The differences between our results with small-town observers and their results with other-race observers make it difficult to apply this same reasoning to explain our results, so what is a reasonable alternative? We suggest that it may be more useful to consider our results in the context of a dual face-space model that entails the construction of separate face-space representations for intra-personal variability and extra-personal variability (Moghaddam, Jebara, & Pentland, 2000). In these models, identification is accomplished by determining the nature of appearance variability that an observer has encountered rather than placing an individual image within an identity-specific region. Within this framework, we suggest that small-town observers differ from large-town observers in that they have a particularly narrow representation of within-person variability that is built primarily from observing the small number of faces in their environment. A consequence of this is that small-town observers may be very hesitant to accept that two images depict the same person, because most instances of face variability will not “fit” well within the distributions they have learned. This would limit observers’ tendency to assign images of different people to the same group and may also contribute to their willingness to create multiple groups of identities when sorting face and body cards. Of course, the extent to which a dual face-space model offers a better account of face recognition ability than a single face-space model is a substantial theoretical issue that requires much more evidence than we can provide here. Still, we offer this interpretation as one theoretical perspective on how our results may arise, and an indication of how continued work with this population may reveal intriguing new directions for future work.

Finally, besides the potential for properties of the input (e.g. face variability) or an underlying model of face representation to influence face recognition abilities in our two groups of observers, it may also be the case that varying properties of the observers themselves may be relevant to consider as well. Face recognition appears to be a highly specific process (Wilmer et al., 2010, 2012) insofar as it does not tend to correlate with other cognitive and perceptual abilities, but there are aspects of observer personality that do appear to predict face recognition abilities. For example, Li et al. (2010) demonstrated that high vs.

low levels of extraversion led to better performance in an old/new face memory task, but did not predict performance with other complex images. More recent results (Lander & Poyarekar, 2015) suggest that extraversion may be more closely associated with famous face recognition in particular, but largely support the proposal that extraverts are more effective at face recognition than introverts. In related work, Davis et al. (2011) and Megreya & Bindemann (2013) reported that lower levels of social anxiety were associated with better face recognition performance. In the latter study, the authors suggest that the effects of personality on face recognition are limited to traits related to neuroticism, which is contrary to the aforementioned results suggesting that extraversion may also influence face recognition ability. Nonetheless, what all of these results have in common is the underlying hypothesis that individual differences in personality may contribute to observers' abilities to recognize the people around them. In terms of our small-town observers, these results suggest that characterizing these aspects of personality in these populations may also be an important means of understanding where variability in face recognition performance originates.

Our current results thus raise a number of interesting questions regarding the manner in which properties of the environment and properties of the observer may give rise to measurable individual differences in face recognition performance. We have demonstrated that a relatively common form of face deprivation (or perhaps hyper-exposure to faces in urban dwellers) leads to decrements in face memory performance, but also appears to improve face recognition when observers are asked to cope with natural variability in face appearance. Unlike other studies in which observers from isolated or depopulated communities have been investigated, our results are not evident at the level of face categories like race, but instead show that individual differences in community size contribute to variation in own-race face recognition. In future work, we hope to both more closely characterize the nature of observers' face environment and to explore the contribution of early experience (face exposure in one's hometown) vs. late experience (arrival in a densely populated environment) on these and other face recognition tasks. If we continue to use effects like the other-race and other-age effects as a sort of model system for predicting how varying amounts of total face experience may impact recognition, there are a number of results that suggest relatively late plasticity of face recognition abilities. For example, in the context of the other-age effect, individual differences in experience at both stages appears to affect the magnitude of the other-age effect for infant faces vs. adult faces (Macchi Cassia et al., 2009), suggesting that experience can impact recognition across a range of time scales. Also, the other-race effect has been shown to be reversible during childhood (Sangrigoli et al., 2005) and at least malleable following training in adulthood (Goldstein & Chance, 1985), which further supports the hypothesis that more generally, individual differences in face exposure are not a fixed property of observers, but may instead have a dynamic impact on visual recognition. Overall, we suggest that our results offer novel insights into the relationship between visual experience and face recognition, and lead to exciting new questions about how observers' abilities reflect the visual world that they are and have been immersed in.

Acknowledgements

This research was supported by NSF grant BCS-1348627, awarded to BB. Thanks to Ganesh Padmanabhan for technical assistance, to Hannah Pearson for permission to use the images in Figure 1, and also to Erin Conwell for assisting with our linear mixed models analysis.

References

1. Andrews S, Jenkins R, Cursiter H, Burton AM. Telling faces together: Learning new faces through exposure to multiple instances. *The Quarterly Journal of Experimental Psychology*. 2015; 68:2041–2050. [PubMed: 25607814]
2. Balas B. Bayesian face recognition and perceptual narrowing. *Developmental Science*. 2012; 15(4): 579–588. [PubMed: 22709406]
3. Balas B. Developing race categories in infancy via Bayesian Face Recognition. *Visual Cognition*. 2013
4. Balas B, Saville A. N170 face specificity and face memory depend on hometown size. *Neuropsychologia*. 2015; 69:211–217. [PubMed: 25659502]
5. Benton, AL., Sivan, AB., Hamsher, KDS., Varney, NR., Spreen, O. Contribution to neuropsychological assessment. Oxford University Press; New York, NY: 1983.
6. Bowles DC, McKone E, Dawel A, Duchaine B, Palermo R, Schmalzl L. Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*. 2009; 26:423–455. [PubMed: 19921582]
7. Bruce V, Henderson Z, Greenwood K, Hancock PJB, Burton AM, Miller P. Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*. 1999; 5:339–360.
8. Bruce V, Henderson Z, Newman C, Burton AM. Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*. 2001; 7:207–218. [PubMed: 11676099]
9. Cassia VM, Kuefner D, Picozzi M, Vescova E. Early experience predicts later plasticity for face processing: Evidence for the reactivation of dormant effect. *Psychological Science*. 2009; 20:853–859. [PubMed: 19493318]
10. Davis JM, McKone E, Dennett H, O'Connor KB, O'Kearney R, Palermo R. Individual Differences in the Ability to Recognise Facial Identity Are Associated with Social Anxiety. *PLoS ONE*. 2011; 6(12):e28800. doi:10.1371/journal.pone.0028800. [PubMed: 22194916]
11. de Heering A, Maurer D. Face memory deficits in patients deprived of early visual input by bilateral congenital cataracts. *Developmental Psychobiology*. 2014; 56:96–108. [PubMed: 23192566]
12. de Heering A, Rossion B, Maurer D. Developmental changes in face recognition during childhood: Evidence from upright and inverted faces. *Cognitive Development*. 2011; 27:17–27.
13. de Heering A, de Liedekerke C, Deboni M, Rossion B. The role of experience during childhood in shaping the other-race effect. *Developmental Science*. 2010; 13:181–187. [PubMed: 20121874]
14. DeBruine LM. Facial resemblance enhances trust. *Proceedings of the Royal Society B: Biological Sciences*. 2002; 269:1307–1312. [PubMed: 12079651]
15. Duchaine B, Nakayama K. The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*. 2006; 44:576–585. [PubMed: 16169565]
16. Dunbar RIM. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*. 1992; 22:469–493.
17. Goldstein AG, Chance JE. Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*. 1985; 23:211–214.
18. Hamilton MJ, Milne BT, Walker RS, Brown JH. Nonlinear scaling space use in human hunter-gatherers. *PNAS*. 2007; 104:4765–4769. [PubMed: 17360598]

19. Jenkins R, White D, van Montford X, Burton AM. Variability in photos of the same face. *Cognition*. 2011; 121:313–323. [PubMed: 21890124]
20. Johnson RA, Edmonds AJ. Familiar and unfamiliar face recognition: A review. *Memory*. 2009; 17:577–596. [PubMed: 19548173]
21. Lander K, Poyarekar S. Famous face recognition, face matching, and extraversion. *The Quarterly Journal of Experimental Psychology*. 2015; 68:1769–1776. [PubMed: 25811985]
22. Laurence S, Zhou X, Mondloch CJ. The flip side of the other-race coin: They all look different to me. *British Journal of Psychology*. 2015
23. Le Grand R, Mondloch CJ, Maurer D, Brent HP. Early visual experience and face processing. *Nature*. 2001; 410:890. [PubMed: 11309606]
24. Le Grand R, Mondloch CJ, Maurer D, Brent HP. Impairment in holistic processing following early visual deprivation. *Psychological Science*. 2004; 15:762–768. [PubMed: 15482448]
25. Li J, Tian M, Fang H, Xu M, Li H, Liu J. Extraversion predicts individual differences in face recognition. *Communicative & Integrative Biology*. 2010; 3:295–298. [PubMed: 20798810]
26. Maurer D, Le Grand R, Mondloch CJ. The many faces of configural processing. *Trends in Cognitive Sciences*. 2002; 6:255–260. [PubMed: 12039607]
27. McCarty C, Killworth PD, Bernard HR, Johnsen E, Shelley G. Comparing Two Methods for Estimating Network Size. *Human Organization*. 2000; 60:28–39.
28. McKone E, Hall A, Pidcock M, Palermo R, Wilkinson RB, Rivolta D, Yovel G, Davis JM, O'Connor KB. Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: Evidence from the Cambridge Face Memory Test – Australian. *Cognitive Neuropsychology*. 2011; 28:109–146. [PubMed: 22122116]
29. Megreya AM, Bindemann M. Individual differences in personality and face identification. *Journal of Cognitive Psychology*. 2013; 25:30–37.
30. Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. *Pattern Recognition*. 2000; 33:1771–1782.
31. Mondloch CJ, Robbins R, Maurer D. Discrimination of facial features by adults, 10-year-olds, and cataract-reversal patients. *Perception*. 2010; 39:184–194. [PubMed: 20402241]
32. Mondloch, CJ., Le Grand, R., Maurer, D. Early visual experience is necessary for the development of some – but not all – aspects of face processing.. In: Pascalis, O., Slater, A., editors. *The development of face processing in infancy and early childhood*. Nova Science Publishers, Inc.; New York: 2003. p. 99-117.
33. Rennels JL, Davis RE. Facial experience during the first year. *Infant Behavior and Development*. 2008; 31:665–678. [PubMed: 18554724]
34. Robbins RA, Nishimura M, Mondloch CJ, Lewis TL, Maurer D. Deficits in sensitivity to spacing after early visual deprivation in humans: A comparison of human faces, monkey faces, and houses. *Developmental Psychobiology*. 2010; 52:775–785. [PubMed: 20564328]
35. Rossion B, Jacques C. Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage*. 2008; 39:1959–1979. [PubMed: 18055223]
36. Sangrigoli S, Pallier C, Argenti AM, Ventura VAG, de Schonen S. Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*. 2005; 16:440–444. [PubMed: 15943669]
37. Scott LS, Monesson A. The origin of biases in face perception. *Psychological Science*. 2009; 20:676–680. [PubMed: 19422630]
38. Sugden NA, Mohamed-Ali MI, Moulson MC. I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology*. 2014; 56:249–261. [PubMed: 24285109]
39. Wilmer JB, Germine L, Chabris CF, Chatterjee G, Gerbasi M, Nakayama K. Capturing specific abilities as a window into human individuality: the example of face recognition. *Cognitive Neuropsychology*. 2012; 29:360–392. [PubMed: 23428079]
40. Wilmer JB, Germine L, Chabris CF, Chatterjee G, Williams M, Loken E. Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*. 2010; 107:5238–5241.

41. Young AW, Hellowell D, Hay DC. Configurational information in face perception. *Perception*. 1987; 16:747–759. [PubMed: 3454432]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

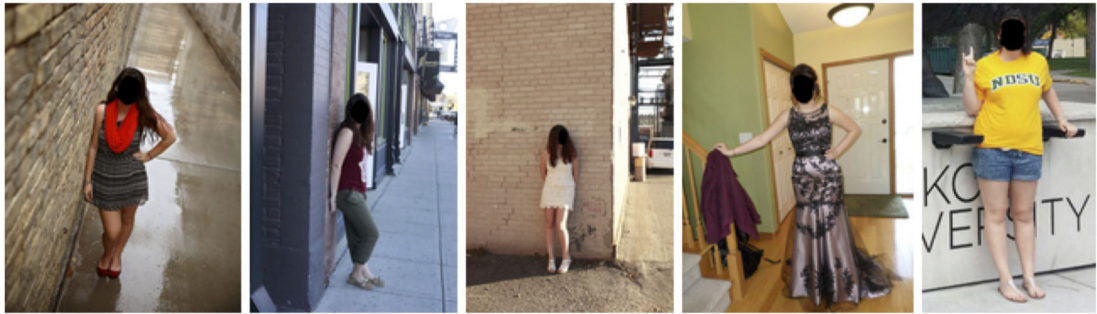


Figure 1.

Due to copyright restrictions, we are unable to display the actual images used for our unconstrained card-sorting task. These images, however, are representative of the appearance of the stimuli and reflect the natural variability in the photographs we chose to use and the cropping procedures used to make our stimuli. These images all depict the same individual.

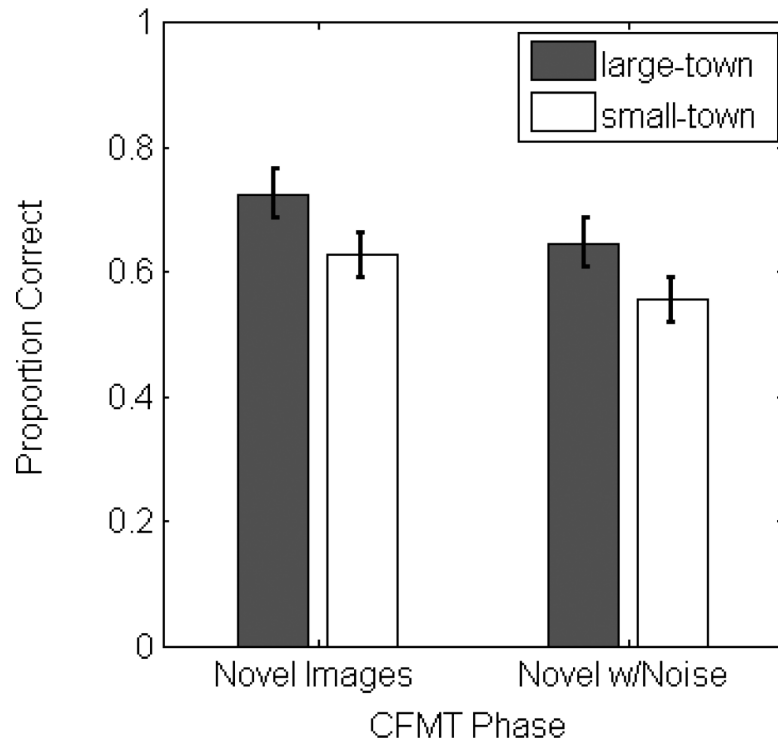


Figure 2. Average CFMT performance in the “Novel Images” and “Novel Images with Noise” phases for small-town and large-town observers. Error bars represent ± 1 s.e.m.

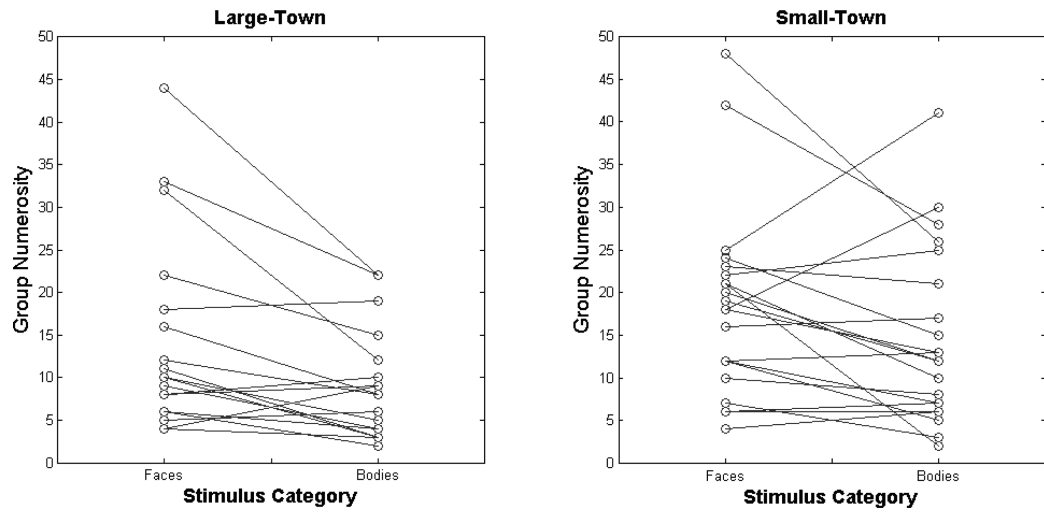


Figure 3.

A plot of the number of groups each participant made in the Face and Body card-sorting task as a function of hometown size. Data points belonging to the same participant are connected by solid lines. Along with Table 1, this figure highlights the variability observed across participants in this task.

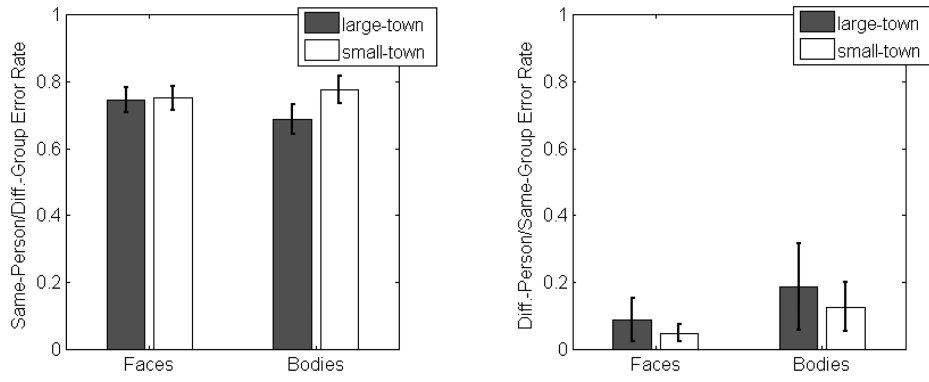


Figure 4. Average error rates (both Same-Person/Different-Group and Different-Person/Same-Group errors) for face and body card-sorting solutions obtained from small-town and large-town participants. Error bars represent ± 1 s.e.m.

Table 1

Median number of groups made by small-town and large-town observers during unconstrained face and body card-sorting. We include the range to convey the high variability of these values across participants.

	Face Groups	Face Range	Body Groups	Body Range
Small-Town	18	[4-48]	12	[2-41]
Large-Town	10	[4-44]	8	[2-22]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Values of Pearson's R and associated p-values for CFMT/sorting error rates across the entire participant sample (N=39). All values reflect two-tailed tests and have not been corrected for multiple comparisons.

Error Type	R	p
Faces (Diff. Person/Same Group)	0.305	0.059
Faces (Same Person/Diff. Group)	-0.540	<0.001
Bodies (Diff. Person/Same Group)	0.329	0.041
Bodies (Same Person/Diff. Group)	-0.269	0.098

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript