Detecting Anomalous Features in Complex Stimuli: The Role of Structured Comparison

Kenneth J. Kurtz
Binghamton University

Dedre Gentner
Northwestern University

The ability to detect anomalies in perceived stimuli is critical to a broad range of cognitive tasks, yet acquiring this ability often requires lengthy practice. In this research, we asked whether findings from research on analogical comparison can be used to aid in the acquisition of perceptual expertise. Building on findings that comparison can facilitate the detection of differences, the present research addressed two questions: (1) Does having an alignable comparison standard improve performance on a difficult detection task? (2) Can such comparison experience improve subsequent detection performance on single anomalous targets? Across 3 experiments, university undergraduates were asked to find an anomalous bone in drawings of animal skeletons. Target items including an anomaly were presented either alone or with a correct standard. Furthermore, to evaluate the impact of ease of alignment, the correct standard was presented either mirror-reversed (low alignable) or regular (high alignable). Results showed increased accuracy when a comparison standard was present and further gains when the standard was more easily alignable. In Experiment 3, we used a between-subjects design to reveal that advance comparison (as opposed to single-item training) led to improved detection of anomalies in subsequent novel examples presented as isolated targets. We conclude that the availability of a standard and ease of alignment promote encoding and processing. Furthermore, comparison-based learning confers an ongoing advantage even without standards for comparison. Therefore, task performance in application areas requiring detection of nonobvious anomalies can be improved by providing alignable standards next to targets or in advance training.

Keywords: structural alignment, comparison, feature detection, perceptual learning

Detecting anomalies in complex perceptual stimuli can be challenging even after extended experience of a domain. In their pioneering work on chicken sexing, Biederman and Shiffrar (1987) found that learning to detect the sex of a newborn chick usually requires extended training and years of exposure to examples. Likewise, Lesgold (1984) documented the intensive training required for radiologists to learn to detect lung spots that could indicate cancer in chest X-rays. Even though the learner knows in advance that differences exist between healthy and diseased lungs, reliably detecting the diagnostic indicators can take years of training and remains difficult even after seeing many examples (Lesgold et al., 1988). One reason for this, as noted by Kok, de Bruin, Robben, and van Merriënboer (2013), is that domain novices often have trouble discriminating relevant from irrelevant information in complex visual tasks, such as diagnosing radiological images, in part because novices focus on features that are intrinsically conspicuous or salient but that may not be relevant to the task (Lowe, 1999).

In view of the wide-ranging importance of anomaly detection in contexts such as medical diagnosis, security, and quality control, better techniques for performing and improving at such tasks would be very useful. In this article, we investigate the role of comparison processes for enhancing people's ability to detect subtle anomalies in complex stimuli. Specifically, we hypothesized that comparison processes can facilitate detection of such anomalies in two ways: (1) by highlighting key differences between the stimulus and the comparison standard and (2) by promoting the acquisition of generalized knowledge of the kinds of structures and anomalies that exist in a domain. There is considerable evidence that comparison processes can highlight certain kinds of differences between items compared. We asked whether this process can be harnessed to promote assessment of perceptual anomalies and whether the results will carry forward to the subsequent assessment of single cases.

We begin by reviewing evidence on the difficulty of learning to detect nonobvious perceptual features and then discuss recent evidence on the process of comparison and its role in difference detection.

Feature Discovery

As introduced above, learning to detect subtle perceptual features can be highly challenging. Such difficulties can be seen not only in expert domains such as radiology but also in familiar domains. For example, in interpreting children's drawings, people do not simply read off features from the stimuli (Wisniewski & Medin, 1994), and even interpreting concrete, observable features requires considerable learning and mediation from top-down processes (see Schyns, Goldstone, & Thibaut, 1998, for a review).

In a notable demonstration, Brooks, LeBlanc, and Norman (2000) asked medical students and expert diagnosticians to describe photographs showing prototypical symptoms of diseases such as pancreatitis. Although the experts identified more features of the diseases than did the students, both groups showed increased sensitivity to features in the photographs when first provided with the correct diagnosis. A particularly surprising aspect of these findings was the widespread failure of participants to notice supposedly *obvious* features. Brooks et al. (2000) noted that this stems in part from the fact that naturalistic stimuli are typically characterized by ambiguity as to what should be taken as a feature. Furthering the case that features cannot be taken for granted, the experience that people have classifying examples influences their ability to segment novel objects into componential part structure (Schyns & Murphy, 1994; Schyns & Rodet, 1997), as well as their psychophysical sensitivity in discrimination performance (Goldstone, 1994). Judgments of perceptual similarity are also systematically affected by learning to classify examples (Goldstone, Lippa, & Shiffrin, 2001; Kurtz, 1996; Livingston, Andews, & Harnad, 1998). In sum, the challenges inherent in noticing, identifying, and evaluating features can be quite serious.

Detecting Anomalous Features and Alignable Differences

The psychological task of detecting anomalous features, that is, errors, faults, or atypical aspects of stimuli, commands interest for a number of reasons. Noticing distinct or unusual features is a critical aspect in a number of practical and applied activities including error detection, pattern classification, troubleshooting, identification, and diagnosis. More generally, such features are likely to weigh heavily in the cognitive tasks of identification, classification, search, discrimination, explanation, and prediction. Anomalous features also play an important role in learning by guiding feature discovery and the development of more sophisticated or better differentiated concepts.

How then are anomalous features noticed and learned? Chun and Jiang (1998) discuss a number of factors that influence visual deployment (see also Kellman, 2002; Wolfe, 1994; Yantis, 1996). These include bottom-up, image-driven factors such as salience (Bravo & Nakayama, 1992; Egeth, Jonides, & Wall, 1972; Theeuwes, 1992; Treisman & Gelade, 1980) and top-down factors such as familiarity (Wang, Cavanagh, & Green, 1994) and expectancy (Loftus & Mackworth, 1978; Miller, 1988; Shaw, 1978; Shaw &

Shaw, 1977). Such factors are clearly important in detecting anomalies, but they cannot be the whole story. Novices may lack sufficient knowledge to make use of top-down expectancies, and, as reviewed above, bottom-up features such as perceptual salience may lead novices to focus on conspicuous features rather than on those relevant to the task (Kok et al., 2013; Lowe, 1999). Our immediate interest lies in those cases in which neither top-down domain familiarity nor bottom-up perceptual salience serves to make the anomaly obvious. We asked whether comparison processes can play a useful role in such cases.

Comparison and Difference Detection

Why should comparison help people detect anomalies? Research on the comparison process helps to explain why juxtaposing examples could be helpful. First, it is well established that comparing two things increases the salience of their commonalities (Boroditsky, 2007; Catrambone & Holyoak, 1989; Gentner & Namy, 1999; Gick & Holyoak, 1983; Loewenstein, Thompson, & Gentner, 1999; Markman & Gentner, 1993b; Tversky, 1977). Less obviously, comparison processes also act to highlight certain specific differences. According to structure-mapping theory, the comparison process first establishes a *structural alignment* between the two situations, a set of correspondences between two structured representations, based on matching the two relational systems (Gentner, 1983, 2003; Gentner & Markman, 1997; Markman & Gentner, 1993a, 2000). (For a description of a computational process model—the Structure-Mapping Engine—see Falkenhainer, Forbus, & Gentner, 1989; Forbus, Gentner, & Law, 1995; Sagi, Gentner, & Lovett, 2012.) The structural alignment process highlights common systems of connected relations (Clement & Gentner, 1991; Gick & Holyoak, 1983) and also *alignable differences*, differences that are connected to the common structure (Gentner & Markman, 1994; Markman & Gentner, 1993a).

An important prediction of structure-mapping theory—and one central to the logic of this article—is that carrying out a comparison increases the salience of alignable differences. This line of prediction is supported by evidence with both perceptual and conceptual materials. First, people find it easier to list differences for high-similarity (alignable) pairs than for low-similarity pairs. For example, Gentner and Markman (1994) gave participants a *speeded-difference* task in which they were given a large set of word pairs and told to list one difference for as many pairs as possible within a time limit. Participants listed 3 times as many differences for high-similarity pairs as for low-similarity pairs, and most of these differences were alignable differences. Alignable differences are typically stated as different values on a common dimension or predicate (e.g., squirrels have fluffy tails, mice have thin tails), whereas nonalignable differences are stated by asserting a fact for one item and denying it for the other (e.g., squirrels have feet, carpets do not). These and other findings suggest that high-similarity pairs can be rapidly aligned, leading participants to notice alignable differences (Gentner & Gunn, 2001; Gentner & Markman, 1994).

Although structure mapping has often been applied to conceptual analogies such as the atom/solar system analogy (Gentner, 1983), there is considerable evidence that the same structural alignment process also occurs during perceptual comparison (Markman & Gentner, 1993b, 1996; Christie & Gentner, 2010;

Kurtz, Boukrina, & Gentner, 2013). In this case, common relational structure is defined by spatial relations, and an alignable difference is one that occupies the same role in the spatial structure of the two items compared. For example, in Figure 1, the central white circle in A versus the central black circle in B constitutes an alignable difference. Nonalignable differences are those that do not occupy corresponding roles, such as the black circle in B versus the lion in C. When asked to name differences, people typically state alignable differences in terms of a common predicate or dimension with contrasting values (e.g., "The center is white in A and black in B"); in contrast, when naming differences for nonalignable pairs (such as B and C), people generally state a feature of one item and deny it for the other: for example, "C has a lion, B doesn't" (Markman & Gentner, 1993a; Sagi et al., 2012). As for conceptual comparisons, the majority of the differences produced are alignable differences. These parallels between perceptual and conceptual comparisons are consistent with the idea that the same process is at work over different kinds of materials.

Response-time studies of visual comparison bear out the claim that structural alignment can influence perceptual availability. When asked to state a difference between two figures, people are faster to name a difference between two high-similarity visual figures (such as A and B in Figure 1) than between two low-similarity figures, such as B and C (Gentner & Sagi, 2006; Lovett, Gentner, Forbus, & Sagi, 2009; Sagi et al., 2012), consistent with the idea that high-similarity pairs are faster to align.

In sum, according to the structure-mapping framework, differences between two compared examples that play the same role within the common relational structure are rendered salient as alignable differences. Thus, in tasks that require anomaly detection, the machinery of comparison can serve to highlight important but initially low-salient features. This suggests a way to make anomalies and other nonobvious features more apparent to learners. The idea is that by comparing a target with a standard, the challenging task of anomaly detection can be converted to the easier task of detecting an alignable difference. If the observer can capitalize on the fruits of the alignment process, then detecting an anomalous feature may become a matter of noticing a difference made salient by comparison.

The above line of reasoning predicts that presenting alignable standards along with the targets will promote detection of anomalous target features that are alignable differences. We tested this prediction using drawings of vertebrate skeletons. Each target item had an anomalous feature—an incorrect bone—and the partici-

pant's task was to click on this bone. Figure 2 shows a sample set with the incorrect bone boxed in the anomalous target (the middle figure).

In designing these studies, our goal was to create a simplified laboratory analog of the kind of challenge that early learners face in complex domains such as radiology or equipment maintenance and troubleshooting. For example, in diagnosing radiographic images (such as chest X-rays), a major challenge is learning to discriminate the relevant information—which is often not of high perceptual salience—from the rest of the complex scene (Kok et al., 2013; Krupinski, 2010; Wood, 1999). There will typically be many irrelevant features, such as the shadow of a nipple, and some important features may be obscured; for example, tumors might be masked by adjacent ribs (Samei, Flynn, Peterson, & Eyler, 2003). Similar challenges of identifying key information from the rest of a complex scene are present in troubleshooting tasks that require detecting an anomaly in a system of relevant and irrelevant interrelated parts, elements, and features. In designing our materials, we avoided salient anomalies that would pop out to the viewer under casual observation, as well as features at the limits of perceptual acuity. We focused on a middle ground consisting of detectable yet nonobvious anomalies embedded in complex objects.

Radiologists examining chest X-rays have to be able to detect a large number of different anomalies that signal different lung diseases (Kok et al., 2013); in many cases, these anomalies can occur at many places within the lungs. A similar variety of different anomalies can occur in troubleshooting contexts in which the fault is likely to reflect misplacement or misalignment of elements or an improper, wrongly sized, broken, distorted, or damaged element. Likewise, in our study, erroneous bones could differ from the correct bone in size, shape, or orientation, and they could occur in different places in a skeleton. In addition, professional tasks based on anomaly detection can require a level of generality of detection expertise. For example, in troubleshooting, this would include the range of products made in a certain manufacturing setting, products made by a particular manufacturer, or products made for a particular purpose or setting.

Another issue we considered is that of time pressure. Tasks such as radiological diagnosis and equipment troubleshooting can be self-paced or there can be time sensitivity. As Kok et al. (2013) noted, in many medical contexts there is de facto pressure on radiology students and residents to act quickly because of the sheer volume of work; likewise, troubleshooting is sometimes done under critical time pressure. Therefore, we carried out studies in both modes. In our first study, participants were given a self-paced task. In the second and third studies, we tested whether comparison processes would remain effective when participants were under time pressure.

Overview of Experiments

The goal of these experiments was to test whether it is possible to capitalize on the psychological salience of alignable differences to support noticing anomalous features in a detection task. We addressed this issue by creating target skeletons, each containing an anomalous bone. Across a series of trials, participants were asked to find and click on the anomalies. The target skeletons were shown either alone, with a high-alignable standard or with a
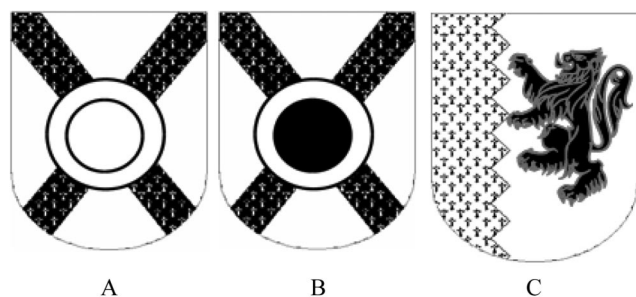
Figure 1. Sample materials used in studies of difference listing (Sagi et al., 2012), showing alignable (A and B) and nonalignable (B and C) pairs.

Low-alignable   Target with   High-alignable
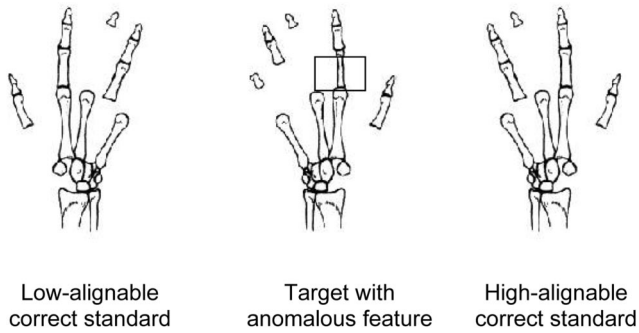correct standard   anomalous feature   correct standard

Figure 2. Sample set of materials used in the present studies, showing the target with an anomalous feature, the high-alignable standard, and the low-alignable standard. The anomaly is shown with surrounding rectangle for illustrative purposes; participants saw the materials without these rectangles.

low-alignable (mirror-reversed) standard, as amplified below (see Figure 2). The low-alignable standard was always the mirror-reverse of the high-alignable standard. Thus, the two standards were equal in terms of the information potentially present, but they differed in their perceptual alignability with the target.

Our prediction based on structure-mapping theory was that presentation of the anomalous target along with an alignable standard would improve the accuracy and efficiency of performance, specifically because aligning the two figures would render the alignable differences salient. A further prediction was that performance would be better with a high-alignable standard than with a low-alignable one. This prediction requires some unpacking. Obviously, the finding that performance improves when a correct standard is present would not tell us what process is involved. Having a correct standard provides information that people could use to improve their accuracy in various ways—for example, to check guesses that they have arrived at by scanning the target. But if the process is one of structural alignment between the standard and the target, then the ease with which people can accomplish this alignment should contribute to their level of performance. Thus, the high-alignment trials should show an advantage over less alignable trials as well as over the solo trials. A further, although more speculative prediction (assessed in Experiment 3), is that comparison of anomalous targets with standards during a study phase would foster learning of the relational structure and thus lead to better performance on further targets.

Experiment 1

The goal of this experiment was to test whether it is possible to capitalize on the psychological salience of alignable differences to support noticing anomalous features in a detection task. We addressed this issue using a within-subjects design to test the accuracy and speed of detecting errors in trials of three types: anomalous target alone, target with high-alignable standard, and target with low-alignable standard. Participants were told that target skeletons had been assembled by student archaeologists. The instructions made clear to participants that the skeletons might be missing some bones, but that their task was to look for any erroneous bone and click on it. Participants were also told that on

some trials they would see a standard assembled by experts. Like the student skeletons, it might have missing bones, but it would have no incorrect bones. Figure 2 shows a sample set and Figure 3 shows the two kinds of paired trials (in single-item trials, the target was seen by itself).

By design, the anomalous features in these stimuli were not rendered salient by strong perceptual magnitude or top-down expectation. Therefore, we expected participants to have difficulty with the task, that is, lengthy response times and high error rates. We expected better performance on trials that offered comparison with a correct standard. Furthermore, if structural alignment processes contribute to performance, we should see better performance with high-alignable than with low-alignable standards.

The low-alignable standard was always the mirror-reverse of the high-alignable standard. Thus, the two standards were equal in the information present, but differed in their perceptual alignability with the target. The alignment account is supported to the extent that we found an advantage for the high-alignable condition over the low-alignable condition.

Method

Participants. The participants in the study were 75 undergraduates taking introductory psychology at Northwestern University. They received partial course credit for their participation.

Materials. The materials were 27 sets of pictures of skeletons taken from elementary science books. Participants were expected to have some limited familiarity in this domain, but to lack sophisticated knowledge or expertise. Although the general domain was not entirely new to the participants, the individual items were novel. The materials were designed to mirror a moderately difficult feature-detection task. To this end, the drawings were detailed illustrations as opposed to schematic representations (see Figures 2–4). The mean number of bones in the target skeletons was 44 (range = 18–120). Each set contained three drawings, designed to allow a within-subject manipulation of presentation condition. The sets consisted of a correct standard, a target identical to the standard except for an anomalous feature, and a mirror-reverse of the standard (i.e., the low-alignable standard). For each target image, one bone was modified to create the critical anomaly by distorting its size, shape, or orientation. All of the anomalies were errors of commission; that is, missing bones did not count as anomalies.

We sought to achieve a stimulus set spanning a reasonably broad range. The materials included 21 mammals, along with three reptiles, two amphibians, and one fish. Two of the mammal images were of human hands. The 27 triplets included seven that were largely symmetric (i.e., frontal views) and 20 that were nonsymmetric. All three pictures in each triplet had some bones missing (not necessarily the same ones) relative to the complete version of the skeleton. In all cases, when a bone corresponding to the modified bone was visible on the opposite side of the skeleton, it was removed from all versions to prevent the use of internal symmetry as a guide. This was to ensure that participants could not discover the misplaced bone in the symmetric skeletons by simply comparing the two sides within the anomalous target skeleton instead of comparing the target with the standard (we recognize, however, that the former is a strategy that might sometimes be used in professional work domains).
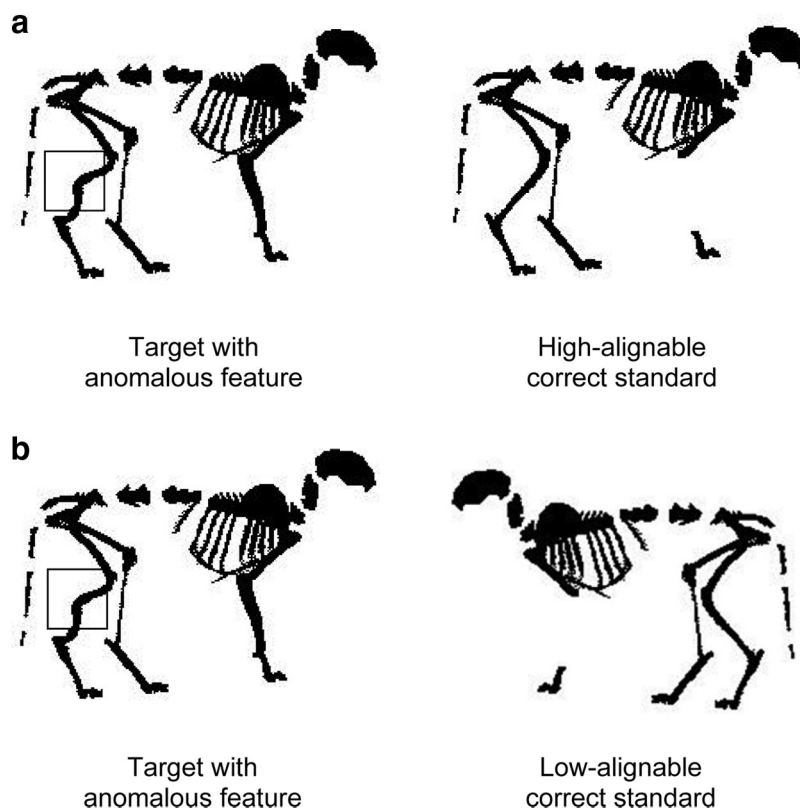
Figure 3. Sample pairs shown in trials, showing target with high-alignable standard (a) and target with low-alignable standard (b). The anomaly is shown with surrounding rectangle for illustrative purposes; participants saw the materials without these rectangles. In single trials, the target was shown alone.

To accommodate the size and shape of the images, we presented eight of the images one above the other and 19 were presented side-by-side. In all cases, the target was displayed on the right or at the top. Figure 4 provides a broader sense of the materials employed including range of difficulty and symmetry.

Calibration. To ensure adequate sensitivity, we needed to verify that the anomalous features of the target were difficult (but ideally not impossible) to notice on their own. We asked seven participants to circle the anomalous bones in the targets. They were given a simplified paper-and-pencil version of the task (no background information on the materials, no practice, no feedback or time pressure) consisting of a packet with each target on a separate page and instructions to circle the anomalous bone. Given the mean number of bones in the figures (44; range = 18–120), chance performance would yield an accuracy of about 2% correct. Participants' success rate was 53%, obviously far greater than chance, but low enough to allow us to observe effects of presenting the standard.

Procedure. There were three within-subjects conditions: single, high-alignable, and low-alignable. In the single condition, trials consisted of the anomalous target only; in the high-alignable condition, the target and the standard were presented together; and in the low-alignable condition, the target and the mirror-reverse of the standard were presented together. Participants received nine items from each condition arranged such that each participant saw only one item from each triplet. Overall, each image set was counterbalanced to occur with equal frequency in all three conditions.

Participants read instructions displayed on a computer screen. The experiment began with a practice session in which participants were acclimated to the motor task of clicking on a location quickly and accurately. They were asked to click with the mouse as quickly as possible on a series of dots that appeared on the screen. Participants were then instructed that they would be looking for "mistakes" in pictures of skeletons. Mistakes were further described as a bone being the wrong size or shape (mistakes in orientation were not explicitly mentioned). Participants were instructed to click on a mistake as soon as they found it. As part of the practice phase, participants performed the detection task on a series of five images of skeletons (of the same general nature as the actual items) presented one at a time. After each practice trial, participants were told whether or not their response was correct, but they were not shown the correct answer. Given the difficulty of the task, this practice procedure was intended to promote task fluency with a minimum influence on subsequent detection ability.

Participants were instructed as follows for the experimental task:

> You will see pictures of fossil skeletons assembled by student archeologists. The student archaeologists were not very careful, so each of the skeletons has one mistake in it.
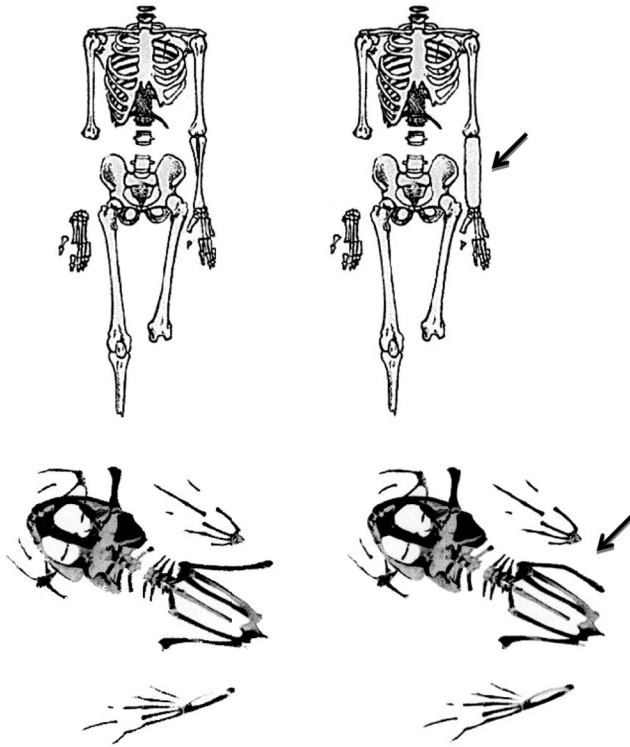
Figure 4. Two examples of stimulus materials showing target and high-alignable standard. The anomaly is indicated with an arrow. The first pair demonstrates a relatively easy example and also illustrates a case of near vertical symmetry. The second pair demonstrates a relatively difficult example.

Your task is to find the mistaken bone and click on it. Please be as quick as possible, but try also to be very accurate. Remember that you only have one chance to click on each picture.

To help you with your task, sometimes you will also see a skeleton marked "Expert Skeleton." This skeleton was put together by expert archaeologists and thus has no mistakes. You can use this to help you spot the mistake in the student skeleton.

The student skeleton will always be marked "Student skeleton: click on the mistake."

Note that you should not be concerned about missing bones. Because these are fossils, all of the skeletons, even the experts' skeletons, are missing some bones. The missing bones do not count as mistakes. The mistakes are bones that are the wrong shape or size.

If you have any questions, please ask them now, before you begin. If not, click on the button below to get started.

After the instructions and practice, each participant was presented with the series of 27 picture sets, nine in each condition as discussed above. Presentation order was randomized for each participant. For each trial, participants had one chance to click on the anomaly, at which point the image was removed. Feedback was given as to whether the response was correct, but the actual correct answer was not provided on incorrect trials.

Results and Discussion

Mean accuracy and mean response time on correct trials were computed across participants. The relatively high error rate suggests that the detection task with minimal time pressure (instructions to respond quickly, but accurately) was quite challenging. In fact, a small subset of participants showed extremely poor performance. Because a basic task competence was needed for useful evaluation of the independent variable, seven of the 75 initial participants were excluded on the basis of failing to correctly answer at least 16 of the 27 trials (error rate greater than 40%). This exclusion threshold was 1 standard deviation from the mean proportion correct. In addition, for 11 of the remaining 68 participants, an individual trial was excluded because of failure to respond in less than 1 min. No two of these excluded trials were from the same participant.

Accuracy. As predicted, accuracy was highest in the two comparison conditions. Mean proportion of correct responding ranged from the least accurate performance ($M = 0.61$) in the single condition to improved performance ($M = 0.81$), in the low-alignable condition to the most accurate performance ($M = 0.87$) in the high-alignable condition (pooled $SD = 0.14$). Performance in the three conditions differed reliably according to a one-way analysis of variance (ANOVA) over mean accuracy, $F(2, 134) = 61.20$, $MSE = 0.02$, $p < .001$. In paired t tests (with alpha level of .05 adjusted using the Bonferroni correction in all tests), each condition was significantly different from every other condition: low-alignable versus single, $t(67) = 6.81$, $p < .001$, $d = 1.43$; high-alignable versus single, $t(67) = 11.61$, $p < .001$, $d = 1.86$; high-alignable versus low-alignable, $t(67) = 2.85$, $p < .01$, $d = 0.43$.

Response time. Participants spent on the order of 10 s to respond to each trial in the detection task. Performance was notably faster in the single condition, and the slowest responding was found in the low-alignable condition. Mean response times for participants' correct responses were 6.76 s in the single condition, 10.65 s in the low-alignable condition, and 9.96 s in the high-alignable condition (pooled $SD = 2.82$). Results of a one-way ANOVA on response time showed a significant effect of condition, $F(2, 134) = 36.69$, $MSE = 7957471.9$, $p < .001$. The difference between the low-alignable and single conditions was significant, $t(67) = 8.691$, $p < .001$, $d = 1.39$, as was the difference between the high-alignable and single conditions, $t(67) = 6.725$, $p < .001$, $d = 1.14$. The high-alignable and low-alignable conditions were not reliably different, $p < .2$.

Considering speed and accuracy together, the single condition was faster than the comparison conditions, but markedly less accurate. At one level, this pattern suggests a speed–accuracy trade-off, whereby the low accuracy in the single condition was the result of participants spending less processing time per item (although note that the low- and high-alignable conditions differed in accuracy, but not in response time). In addition, there is the obvious difference that the single condition provided less information to process. We believe that another level of interpretation is relevant: When in the single condition, participants may have been more likely to give up and respond quickly (but somewhat haphazardly) because they lacked a clear sense of which features were anomalous. In contrast, when in the low- or high-alignable conditions, participants had access to the standard to guide their

processing of the target. The data also suggest that the best combination of speed and accuracy arose in the high-alignable condition, but further evidence is needed to support this claim.

Item effects. There were differences in the level of detection accuracy across items, but no systematic basis was found for the item differences. Most important, the comparison advantage remained evident in each of the three subsets when the items were divided into thirds based on difficulty (mean performance on the detection task). To be sure that highly transparent or highly difficult items did not distort the performance across conditions, we also conducted an analysis with the six easiest and six most difficult items excluded. The results showed the same set of statistically significant differences in accuracy as in the overall analysis. The high-alignable condition ($M = 0.88$) was more accurate than either the single condition ($M = 0.64$) or the low-alignable condition ($M = 0.82$). A notable change in the pattern of results for these moderate-difficulty items (relative to the overall analysis) was that response times were significantly shorter for the high-alignable condition ($M = 9.174$ s) than for the low-alignable condition ($M = 11.278$ s), $t(67) = 2.87$, $p = .005$. Because the available information was equal for these two conditions, this is consistent with our claim that alignability influences how readily the information can be extracted.

These results demonstrate that comparison can act as an aid to perceptual processing. However, one potential concern is whether the anomalies in our studies were simply too obscure to detect without some sort of external guidance. The present findings do not support this interpretation. First, as noted above, the calibration study using only the targets showed above-chance (53%) detection of the anomalies, and the detection rate for anomalies in solo targets was even higher (66%) in Experiment 1. (This could suggest that participants were learning something from the aligned pairs in this within-subject design, a possibility we investigated directly in Experiment 3.) The comparison advantage was observed across the range of item difficulty (see Table 1). Importantly, the comparison benefit was observed even for the easiest items—those for which the anomaly was most clearly evident without a standard.

Lastly, a potential concern arises (we thank an anonymous reviewer for pointing this out) when images were presented side-by-side: The mirror-reversal leads to a difference in the physical distance between the anomaly and its correct version. To ascertain whether the observed condition differences were strictly attributable to alignability, we conducted a follow-up analysis. For each of the side-by-side items, we determined whether the anomaly in the target item was to the right or left of center. Given that the target was displayed to the right of the standard, if the anomaly was right of center (10 cases), this means that the low-alignable presentation (mirror-reversal) put the anomaly closer in physical distance to the

corresponding location in the standard; accordingly, if the anomaly was left of center (nine cases), this means that the high-alignable presentation put the anomaly closer in physical distance to the corresponding location in the standard. In sum, half of the time the high-alignable case was favored and half of the time the low-alignable case was favored by the physical distance.

Summary. The results conformed to the expected pattern in many, although not all, respects. As predicted, we found an advantage for comparison: Participants showed better accuracy in detecting the anomalous feature when given a standard against which to compare the target item than when given only the target. A key aim of this research was to test the processing hypothesis that structural alignment would render alignable differences salient. This hypothesis predicted an advantage not only for the two comparison conditions over the single condition, but more specifically for the high-alignable pairs over the low-alignable pairs. Indeed, the accuracy patterns indicate an advantage based on alignability: High alignability > low alignability > single across the stimuli. However, the response time data show only an advantage for the single condition relative to the two comparison conditions. This pattern is consistent with the alignment account, but also with weaker explanations based on the availability of more information. Even so, there were indications that participants needed less processing time to achieve this accuracy advantage for comparison when the standard was easy to align with the test figures. For example, the response time results for moderate-difficulty items showed an advantage for the high-alignable over the low-alignable condition. In the next experiment, we used a deadline task to test for effects of alignability on the accuracy of anomaly detection.

Experiment 2

Processing complex stimuli for medical diagnosis or fault detection occurs in real time. As discussed earlier, there is often a degree of pressure to respond quickly and move on to other tasks. In such cases, it is desirable to have a high level of *efficiency*, by which we mean high accuracy coupled with rapid performance. In our first experiment, participants were encouraged to work quickly, but they were permitted to take as long as they wanted to respond. In Experiment 2, we used a deadline task to approximate the kind of time-pressured task that can occur under naturalistic conditions. To achieve the desired level of challenge, we set the deadline at 1 standard deviation greater than the mean overall response time in Experiment 1. The prediction from structure-mapping theory is that the relatively rapid alignment process for the high-alignable standard–target pair will result in rapid noticing of the alignable difference (the anomaly). Hence, the high-alignable standard will better promote successful detection than the low-alignable standard or no standard at all. The ease of alignment should contribute not only to higher accuracy, but also to more efficient processing.

Method

Participants. The participants were 58 undergraduates taking introductory psychology at Northwestern University who received partial course credit.

Materials. The materials were identical to those used in Experiment 1.

Table 1
Results of Experiment 1: Mean Accuracy by Condition for Three Levels of Item Difficulty

Item	Single	Low-alignable	High-alignable
Easiest	0.80	0.92	0.94
Moderate	0.59	0.84	0.93
Difficult	0.42	0.67	0.73

Procedure. The design and procedure were the same as those used in Experiment 1 except for the addition of the response deadline. After 13 s without a response, the stimulus was removed and feedback indicating an incorrect response was provided. Participants were instructed about the deadline procedure as follows:

> Your task is to find the mistaken bone and click on it. You will have 13 seconds to find the mistake, so work quickly, but keep in mind that you only have one chance to click on each picture.

Results and Discussion

Mean accuracy was computed as in Experiment 1, with the additional provision that failure to respond before the allotted time elapsed was treated as an error. As expected, detection accuracy dropped sharply under deadline. Ten participants who failed to answer at least 11 of 27 trials correctly were excluded from the analysis (resulting in $n = 48$). (The criterion for exclusion was more lenient than in Experiment 1 because of the increased difficulty of the deadline task.) As in Experiment 1, performance was least accurate ($M = 0.52$) in the single condition, intermediate ($M = 0.57$) in the low-alignable condition, and most accurate ($M = 0.66$) in the high-alignable condition (pooled $SD = 0.17$). A one-way ANOVA on mean accuracy showed a significant difference among conditions, $F(2, 94) = 7.42$, $MSE = 0.03$, $p = .001$. Based on planned comparisons, the difference between the high-alignable and single conditions was significant, $t(47) = 3.89$, $p < .001$, $d = 0.82$, as was the difference between the high-alignable and low-alignable conditions, $t(47) = 2.64$, $p = .01$, $d = 0.53$. The difference between the single and low-alignable conditions was not significant, $p = .26$.

These results confirm our efficiency predictions and bear out the importance of alignability in achieving greater accuracy under conditions of time pressure. If the effects of comparison were simply due to the additional information available through having a correct standard, we would have seen a difference between the comparison conditions and the single condition, but not between the high- and low-alignable conditions. Instead, we found a difference between the high-alignable condition and the rest. In fact, participants showed a comparison advantage under deadline only with an easily aligned standard: The low-alignable condition did not differ significantly from the single condition. These results are evidence for the role of alignment processes in this task. On this account, the advantage of comparison is that by aligning the target with the standard, the anomaly stands out as an alignable difference. The present results buttress the findings of Experiment 1 to show that detection performance is faster and easier with comparison with a high-alignable standard.

Experiment 3

The evidence thus far has shown that comparison can increase the accuracy of anomaly detection, especially when the standard and target are readily alignable. Essentially, by converting an anomaly-detection task into a difference-detection task, we can capitalize on the fact that detecting an alignable difference in structurally aligned pairs is extremely rapid (Gentner & Sagi, 2006; Lovett et al., 2009; Sagi et al., 2012).

However, in terms of professional application, one might be concerned that providing alignable standards could lead to a de-

pendence on such standards. If so, then this technique might not extend to or could even hamper future ability to detect errors without such external aids. But there is another possibility: Comparing cases may promote future anomaly detection, even for single cases, by improving learners' mental models of the materials. There is considerable prior evidence that comparison of structured cases promotes relational encoding and schema abstraction (Catrambone & Holyoak, 1989; Gentner, Loewenstein, & Thompson, 2003; Gentner, Loewenstein, Thompson, & Forbus, 2009; Gick & Holyoak, 1983; Kurtz et al., 2013; Loewenstein et al., 1999; see also Guo, Pang, Fang, & Ding, 2012).

Developmental research indicates that comparison promotes relational learning in children as well as in adults (Childers & Paik, 2009; Christie & Gentner, 2010; Gentner, Anggoro, & Klibanoff, 2011; Gentner & Namy, 1999; Haryu, Imai, & Okada, 2011; Namy & Gentner, 2002). For example, Gentner, Loewenstein, and Hung (2007) taught 3-, 4-, and 5-year-old children to identify novel object parts; children were shown a novel standard figure, told that "this one has a blicket," and asked to say which of two other figures also had a "blicket." Mirroring our findings in Experiments 1 and 2, children performed much better when the alternatives were highly similar to the standard (and thus easily aligned with it) than when they were less similar (and less easily aligned) to the standard. Perhaps more surprising, the results also showed a *progressive alignment* effect: Children who were given highly alignable materials subsequently did better on the low-alignable materials than children who received the same number of trials, all with the low-alignable materials. This progressive alignment pattern—whereby experience with highly alignable pairs potentiates performance on less alignable pairs that have the same relational structure—has been found in several studies (e.g., Gentner et al., 2011; Haryu et al., 2011; Kotovsky & Gentner, 1996; Thompson & Opfer, 2010). It suggests that achieving a structural alignment—which is relatively easy for highly alignable pairs—renders the common relational structure more salient and more available for future use, facilitating transfer to less readily alignable pairs.

These findings motivate an intriguing prediction tested in the present experiment: Comparing anomalous targets with high-alignable correct standards will facilitate detection performance in subsequent test trials with solo anomalous target presentation, perhaps even for novel exemplars. In other words, we asked whether there are benefits of prior comparison even when the informative standard is no longer possible. Obviously, such a finding could have important implications for training in diagnosis and fault detection. In addition, it provides a further test of whether the process of structural alignment between the target and the standard confers benefits over and above the simple presence of extra information (i.e., the correct standard).

The current experiment differed from the previous studies in that it was designed to evaluate the impact of advance comparison. In the new comparison condition, participants studied high-alignable target–standard pairs. In the new version of the single condition, participants studied the correct standards, presented singly, and received each standard twice, thereby receiving twice the number of trials as the comparison group. In this way, both groups received an equal number of overall item exposures. The key to the design was providing the comparison group with an opportunity to perform structural alignment between standards and anomalous targets. The single group was given twice as many

exposures to the correct standards and twice as much overall study time, but no opportunity for direct comparison. If the comparison group showed better performance on the subsequent test than those who had received more experience with correct standards, this would be evidence for the power of comparison in learning.

Method

Participants. The participants in the study were 84 undergraduates taking introductory psychology at Northwestern University. They received partial course credit for participating. Participants were randomly assigned to one of the two conditions.

Materials. The materials for the study and test sets were randomly selected subsets of the materials used in Experiments 1 and 2.

Procedure. Unlike the previous experiments, the current procedure employed a between-subjects design and three distinct phases for each participant (see Table 2).

Study phase. During the study phase, participants simply studied the materials presented. In the comparison condition, 13 alignable pairs were displayed in a random order, one pair at a time, for 10 s each. In the single condition, each of the 13 expert standards was shown twice, one at a time, for 10 s per presentation. All examples appeared once in a random order before being repeated in a new random order. Thus, the two groups were equated in terms of total number of figures seen. The total study time was twice as great in the single condition as in the comparison condition. The most important difference, however, was that the single group had high exposure to the correct skeletons, but the comparison group received high-alignable pairs of correct and incorrect skeletons.

As in the prior studies, both groups were told that, in the main task, they would see pictures of student-assembled skeletons, some of which had mistakes, and that they were to point out the mistakes as quickly as possible by clicking on them. The single group was told that to help them with this task, they would first see skeletons put together by expert archeologists with no mistakes. The comparison group was told that to help them with the task, they would first see pairs of skeletons: one marked "Expert skeleton" that had been put together by an expert archeologist and had no mistakes, and one marked "Student skeleton" that had been put together by a student archeologist and had one mistake. Both groups were told that they did not need to make any response, and that they should "Use this chance to study the skeletons/skeleton pairs since this may help you spot mistakes later in other skeletons."

Transfer test phase. After the study phase, all participants were tested on 10 novel single-item anomalous targets that neither group had seen before. As in the previous studies, participants indicated the location of the anomalous feature by clicking on it. In contrast to the prior studies, no feedback was provided to participants regarding the accuracy of their response. A response deadline of 10 s was used. This was decreased from the prior deadline of 13 s in Experiment 2 in recognition of the fact that the participants were now entering the task after having completed a study phase and to minimize the opportunity for learning during test. The transfer trials were identical across the two conditions.

Retest phase. After a short break and reiteration of task instructions, participants were retested on the same 10 transfer items intermixed with the 13 target skeletons from the study phase. The primary purpose of this phase was to provide a retest of any group differences found with the 10 transfer items. For completeness, we also tested the degree to which participants were able to spot the errors in the initial 13 items (now presented as single items). We expected the comparison group to have an advantage here, as they had seen the anomalous targets paired with the correct standards during study. Thus, the results would only be noteworthy if this advantage failed to appear.

Results and Discussion

The results bear out the predicted advantage of advance comparison. On the key transfer test with novel items, comparison learners ($M = 0.27$) were significantly more accurate at detecting the anomalies than were single learners ($M = 0.14$), $t(82) = 3.75$, $p < .001$, $d = 0.86$ (see Table 3). Because neither group had seen these items, the comparison advantage here is evidence that comparison experience can confer general insight beyond the particular examples studied.

On the retest, the comparison advantage was maintained on the transfer items: Comparison learners ($M = 0.36$) were significantly more accurate than single learners ($M = 0.25$), $t(82) = 2.76$, $p = .007$, $d = 0.61$. As expected, on the 13 old (initially studied) items, the comparison learners ($M = 0.52$) outperformed single learners ($M = 0.37$), $t(82) = 3.23$, $p = .002$, $d = 0.71$. This was not surprising given that the comparison group had had prior exposure to the studied items and the opportunity to compare them with expert versions (although the anomalies themselves were not indicated in the training). Also, as expected, both groups showed better performance on the studied items than on the new items.

The key finding concerns the transfer items. Participants who compared anomalous targets with correct standards outperformed participants who received twice the amount of exposure to the correct standards. This effect extends our earlier findings in an important way: Participants benefited from comparing studied items such that they performed better on test cases that were novel and for which there had been no comparison opportunity. We conclude that comparison processing can promote subsequent performance in a fault-detection task. This means that comparison experience confers a level of insight that is more general than the particular items.

These findings are consistent with the progressive alignment proposal: Structural alignment promotes encoding of structural commonalities, and this informs further perceptual processing in the domain. In this case, the alignment is one of common spatial relational structure (cf. Gattis, 2002, 2004; Markman & Gentner, 1993b, 1996; Wisniewski & Middleton, 2002). Given prior findings that aligning two instances of the same relational structure tends to increase the salience of that structure in subsequent

Table 2
Design of Experiment 3

Phase	Single condition	Comparison condition
Study	2 passes of 13 standards	1 pass of 13 target–standard pairs
Test	10 novel targets; all item retest	10 novel targets; all item retest

Table 3
*Results of Experiment 3: Mean (*SD*) Proportion Correct*

Item	Single	Comparison
Transfer (10)	0.14 (0.13)	0.27 (0.17)**
Studied (13)	0.37 (0.21)	0.52 (0.21)**
Transfer retested (10)	0.25 (0.17)	0.36 (0.19)**

** $p < .01$.

processing (e.g., Gentner et al., 2011; Kotovsky & Gentner, 1996; Markman & Gentner, 1993a), we would expect that the kinds of skeletal configurations that are highlighted as in previous comparisons would be more likely to be picked out in new cases. Such effects may be similar to chess experts' use of their knowledge of allowable spatial configurations in chess to quickly and accurately identify violations (Chase & Simon, 1973).

A further possible contributor to the comparison effect may be learning about the general kinds of anomalies present in the targets. Although comparison and single learners saw the same number of images during training, only the comparison learners saw the anomalous target items. Although the errors were not marked, we suggest that comparison with the standards allowed this group to notice the kinds of errors that need to be detected and to generalize this understanding to novel items and anomalies. It seems likely that both gaining sensitivity to the relational configurations characteristic of a domain and learning about likely anomalies contribute to gaining perceptual expertise in complex domains.

Our findings in Experiment 3 are consistent with findings of Kok et al. (2013), who tested whether structural alignment could help students learn to diagnose diseases from radiographs (X-rays) of the chest. In their study, third-year medical students studied pairs of chest radiographs showing 12 different diseases of the heart and lungs, each labeled as to the disease. In the pathology/ normal condition, a radiograph of a patient was shown next to one of a healthy person; in the pathology/pathology condition, both radiographs showed the same disease. When the students were subsequently asked to diagnose single novel chest radiographs, the pathology/normal group ($M = 0.63$) outperformed the pathology/ pathology group ($M = 0.54$) on focal diseases such as lung tumor (although not on diffuse diseases such as cystic fibrosis, for which both groups performed well). Kok et al. interpret these findings in terms of structural alignment theory: "The normal anatomy on both the normal image and the pathological image can be aligned to each other. The disease-related information, which signifies the main difference between the two images, will then become salient" (p. 2). Indeed, the Kok et al. pathology/normal pairs are quite analogous to our high-alignable pairs (anomalous target/expert standard). Thus, the high performance in this condition serves to generalize the claim that structural alignment processes can aid in learning relevant perceptual regularities.

However, their findings also raise new questions. Their control condition was pathology/pathology pairs showing the same disease. Kok et al. (2013) suggest that this condition was less effective because it lacked the key information needed—namely, the difference between normal and pathological cases. Yet, one might have expected that comparing two instances of the same disease would lead to extracting the common patterns characteristic of the

disease based on the many findings in which comparison had led to increased ability to perceive the common structure in later examples (e.g., Christie & Gentner, 2010; Gentner & Namy, 1999; Gick & Holyoak, 1983; Graham, Namy, Gentner, & Meagher, 2010; Kotovsky & Gentner, 1996; Kurtz et al., 2013). Because there was no single-item control group, it remains possible that both comparison groups improved in varying degrees.

Kok et al. (2013) did not test whether comparing radiographic pairs could support online sensitivity to anomalies (as in our Experiments 1 and 2), but this prediction is consistent with their work. Overall, their results dovetail with our findings in Experiment 3 in suggesting that structural alignment can support learning perceptual structure in complex domains. These findings add to evidence that comparison entails an alignment of common structure that renders common relational structure and alignable differences more salient in current (and future) processing.

General Discussion

Our goal in this research was to test whether analogical comparison could be used to help make anomalous features more detectable. According to structure-mapping theory, comparison with an alignable standard can render anomalous features more detectable by revealing them as differences (between target and standard) that are linked to common structure. Based on this account, we predicted the following effects on task performance: (a) more accurate detection of an anomalous feature when the target could be compared with a standard; (b) more specifically, greater accuracy in detection when given a readily alignable standard than when given one that is harder to align. These two predictions were tested in Experiments 1 and 2. In Experiment 3, we tested a further prediction: (c) Comparison to a standard would lead to highlighting structural commonalities and facilitate subsequent processing of items in the domain.

In Experiments 1 and 2, participants had to identify errors in a target skeleton, which were seen either singly, with a high-alignable standard, or with a low-alignable standard. We found reliable accuracy differences in the predicted direction between all three conditions, including the critical prediction of greater accuracy in high- than in low-alignable conditions. The two comparison conditions did not differ in overall response time (both took longer to process than single-item presentation, which, however, was far less accurate). However, analyses of response latency for the moderate-difficulty items suggested that alignability decreased the time cost of comparison. In Experiment 2, we used a deadline task to test the prediction that high-alignable comparisons would increase the efficiency of online anomaly detection. We found that participants were more accurate under time pressure in the high-alignable condition than in the low-alignable or single condition.

The finding that participants are more accurate at detecting an anomalous feature in a skeleton when they have a comparison standard is consistent with prior findings that analogical comparison highlights commonalities and differences (Tversky, 1977). Furthermore, because the low-alignable pairs differed from the high-alignable pairs only in alignability and not in the information present, these findings support the claim that structural alignment is the critical process here.

Experiment 3 tested the third prediction: Comparison can improve the ability to detect anomalies in future items. Participants

who compared anomalous targets with alignable standards performed better in a subsequent single-item test than did those who received twice as many study exposures to the standards. Importantly, the comparison group was more accurate than the single-item group both on the studied targets and on new targets not previously seen by either group. The finding that comparison with a standard can facilitate subsequent processing even of novel targets is consistent with research in analogy showing that structural alignment highlights common relational structure—including perceptual structure—thereby facilitating future processing of items in the domain (Gentner & Gunn, 2001; Gentner et al., 2007; Gick & Holyoak, 1983; Kotovsky & Gentner, 1996).

These findings join other recent work in which structural alignment is used as a means to an end, such as the Kok et al. (2013) study showing that medical students who compared radiograms of diseased versus healthy people were subsequently better able to diagnose focal lung diseases than a control group. These studies provide evidence that alignment effects are not confined to traditional laboratory tasks, but can also be used in professional work contexts—both in promoting accurate detection of key features and in learning to detect such features in the future.

These results are consistent with prior findings that structural alignment can lead to rapid detection of specific differences (Gentner et al., 2009; Gentner & Sagi, 2006; Lovett et al., 2009; Sagi et al., 2012). However, as far as we know, this study is the first to show that the highlighting of alignable differences observed in comparison processing can be harnessed to aid in more accurate online detection of faults. The fact that alignability aids processing under speeded conditions further suggests that it could be useful in a variety of situations. Finally, our finding that comparison experience can lead to more effective perceptual processing of novel items within the same domain suggests a role for comparison in perceptual learning.

Limitations

Our goal in this research was to test whether structural alignment processes could be used to support fault detection and diagnosis in complex perceptual tasks such as radiographic analysis. To this end, we created low-salient anomalies that were hard to detect purely perceptually and embedded them in complex skeletal backgrounds, mirroring the challenges of detecting symptoms of trauma or disease in chest X-rays. The anomalous bones could be of different shapes, sizes, or orientations from the correct one, and could occur at various places within the skeleton.

However, there are some ways in which our task deviates from the task in the radiography reference situation. First, we told our participants that each student skeleton had exactly one wrong bone; this is clearly a simpler situation than that faced by a radiographer, who must diagnose whether there is disease at all and whether there are multiple problems. Second, our (high-alignable) normal skeletons were always an exact match for the target skeletons except for the anomaly. Although these simplifications almost certainly acted to make the task easier for our participants, we believe that the basic findings will still apply in professional tasks. For one thing, radiologists may compare a person's chest X-rays from two different dates, or they may compare two lungs from the same individual. Practices that make comparison easier should facilitate anomaly detection.

Along these lines, there is a larger question about the generality of the use of alignment to promote anomaly detection. The existing literature (cited above as background on structural alignment) makes clear the wide range of materials with which comparison effects are found. We can reasonably expect the advantage of constructing alignable comparisons to be widely applicable, but there are limits. The characteristics that ought to be in place are the presence of relational structure (be it spatial, conceptual, or a combination), anomalies that are distinct elements (objects, predicates) within such a relational structure, and a substantial degree of commonality in the relational structure across examples within the domain. Fortunately, to a great extent, domains hold their status as domains because they tend to meet these preconditions. However, comparison should be less effective in detection tasks when anomalies are not identifiable elements (such as a diffuse presence, as in Kok et al., 2013) or they are not connected to common structure. In addition, comparison might not prove effective in transfer when there is a lack of coherent structure within examples or a lack of regularity of structure across examples.

Lastly, we note that further experimentation will be needed to determine exactly what participants are learning in these studies. They might be learning schemas for skeletal structure (as we suggest), or learning the set of likely error types, or developing task-specific comparison strategies, or some combination of these. Future work may also tell us which of these best translates to reference situations such as radiographic analysis.

Implications for Perceptual Categorization

As Kellman (2002) noted in his review of perceptual learning, the idea that perceptual learning is facilitated by comparison dates back as least as far as Pavlov. The main focus in this research tradition has been on the use of comparison as a way to differentiate categories; the idea is that the presentation of contrasting items allows learners to learn diagnostic features. Our research in Experiments 1 and 2 extends this tradition to the idea of differentiating various kinds of faults from their normal counterparts. However, the analogical account also suggests that comparison within categories can be valuable in highlighting the relational structure that characterizes the category. The results of Experiment 3, in which aligning pairs of skeletons led to superior performance on skeletons not seen before, are consistent with this kind of structural abstraction.

Extrapolating from these findings, we suggest a much broader role for comparison processes in the learning of perceptual categories. Researchers studying category learning have convincingly argued that the environment alone does not determine the feature space for conceptual representation and that further constraints are needed to determine the features and relations that enter into categorization processes (Murphy & Medin, 1985; Schyns et al., 1998; Wisniewski & Medin, 1994; see also Kurtz & Dietrich, 2013). Developmental evidence suggests that comparison may act to alter representations during category learning by highlighting structural commonalities and differences (Christie & Gentner, 2010; Gentner & Medina, 1998; Kotovsky & Gentner, 1996). Paired presentation of examples has been shown to facilitate learning of novel perceptual relational categories in adults (Goldstone & Son, 2005; Kurtz et al., 2013).

Implications for the Development of Perceptual Expertise

A core implication of this research is that comparison processes can be used to improve performance on difficult perceptual tasks and to accelerate learning of such tasks. The ultimate goal in such learning is the achievement of expert performance. We suggest that one reason that perceptual expertise requires lengthy practice (Biederman & Shiffrar, 1987; Lesgold, 1984; Lesgold et al., 1988) is that becoming an expert requires learning to represent the problem domain in appropriate ways (Brooks et al., 2000; Chi, Feltovich, & Glaser, 1981; Kok et al., 2013; Wood, 1999), to see things differently in the domain. This involves learning to perceive the relational patterns against which particular features—including anomalous features—stand out. For example, Kundel and La Follette (1972) suggest that skilled expert interpretation of radiological images depends not only on knowledge of normal and abnormal features but also on the backgrounds in which they occur, and Klein and Hoffman (1993) state that a major difference between novices and experts is that even when novices perceive all the relevant details, they fail to see the relevant relations among them. This pattern is also seen in children's learning (Gentner & Rattermann, 1991; Halford, 1992). For example, Chipman and Mendelson (1975) studied the development of sensitivity to visual structure and found that it extends over many years and is characterized by an increase in the number of pattern elements that are perceived to be organized. Thus, one reason that becoming an expert takes time is that people must learn which relations to attend to and how these translate into relevant decisions. This is often a lengthy process.

Given the evidence that comparison can lead learners to adopt different encodings—particularly relational encodings—we suggest that comparison-driven learning could be important in acquiring perceptual expertise in professional domains. Going further, structure-mapping theory could be used to develop design principles for using comparison to accelerate learning in complex pattern-detection tasks. Borrowing techniques from developmental studies, training sequences could be designed to use progressive alignment, beginning with highly alignable pairs and progressing through less alignable pairs (Gentner et al., 2007, 2011; Haryu et al., 2011; Kotovsky & Gentner, 1996; Thompson & Opfer, 2010). Other possibilities could also be investigated, such as permitting learners to store items (such as ideal exemplars or borderline cases) that they wish to use in future comparisons.

Conclusion

Noticing anomalous features is critical to a number of professional applications and can be a gateway to understanding and learning, but it is a complex and difficult task. Comparison of an anomalous target with a standard is a valuable tool for promoting the discovery of such features. These findings bear on the basic cognitive processes by which people come to notice previously undetected features. For professional tasks, our findings suggest that providing an alignable comparison standard can facilitate online anomaly detection; equally important, such experience can transfer to future examples in the domain. The study of comparison processes in the laboratory has progressed rapidly in the past decade. Although there remain many open questions, we suggest that perceptual comparison can be a significant accelerator in the acquisition of perceptual expertise.

References

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 640–645. doi:10.1037/0278-7393.13.4.640

Boroditsky, L. (2007). Comparison and the development of knowledge. *Cognition, 102,* 118–128. doi:10.1016/j.cognition.2002.08.001

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics, 51,* 465–472. doi:10.3758/BF03211642

Brooks, L. R., LeBlanc, V. R., & Norman, G. R. (2000). On the difficulty of noticing obvious features in patient appearance. *Psychological Science, 11,* 112–117. doi:10.1111/1467-9280.00225

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1147–1156. doi:10.1037/0278-7393.15.6.1147

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81. doi:10.1016/0010-0285(73)90004-2

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152. doi:10.1207/s15516709cog0502_2

Childers, J. B., & Paik, J. H. (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language, 36,* 201–224. doi:10.1017/S0305000908008891

Chipman, S. F., & Mendelson, M. J. (1975). The development of sensitivity to visual structure. *Journal of Experimental Child Psychology, 20,* 411–429. doi:10.1016/0022-0965(75)90116-2

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

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71. doi:10.1006/cogp.1998.0681

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science, 15,* 89–132. doi:10.1207/s15516709cog1501_3

Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multi-element displays. *Cognitive Psychology, 3,* 674–698. doi:10.1016/0010-0285(72)90026-6

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41,* 1–63. doi:10.1016/0004-3702(89)90077-5

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19,* 141–205. doi:10.1207/s15516709cog1902_1

Gattis, M. (2002). Structure mapping in spatial reasoning. *Cognitive Development, 17,* 1157–1183. doi:10.1016/S0885-2014(02)00095-3

Gattis, M. (2004). Mapping relational structure in spatial reasoning. *Cognitive Science, 28,* 589–610. doi:10.1207/s15516709cog2804_4

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170. doi:10.1207/s15516709cog0702_3

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 195–235). Cambridge, MA: MIT Press.

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure-mapping and relational language support children's learning of relational categories. *Child Development, 82,* 1173–1188. doi:10.1111/j.1467-8624.2011.01599.x

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition, 29,* 565–577. doi:10.3758/BF03200458

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8,* 285–307. doi:10.1080/15248370701446434

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95,* 393–408. doi:10.1037/0022-0663.95.2.393

Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. (2009). Reviving inert knowledge: Analogical abstraction supports relational retrieval of past events. *Cognitive Science, 33,* 1343–1382. doi:10.1111/j.1551-6709.2009.01070.x

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science, 5,* 152–158. doi:10.1111/j.1467-9280.1994.tb00652.x

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52,* 45–56. doi:10.1037/0003-066X.52.1.45

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65,* 263–297. doi:10.1016/S0010-0277(98)00002-X

Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14,* 487–513. doi:10.1016/S0885-2014(99)00016-7

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London, UK: Cambridge University Press. doi:10.1017/CBO9780511983689.008

Gentner, D., & Sagi, E. (2006). Does "different" imply a difference? A comparison of two tasks. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 261–266). Mahwah, NJ: Erlbaum.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38. doi:10.1016/0010-0285(83)90002-6

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123,* 178–200. doi:10.1037/0096-3445.123.2.178

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition, 78,* 27–43. doi:10.1016/S0010-0277(00)00099-8

Goldstone, R. L., & Son, J. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences, 14,* 69–110. doi:10.1207/s15327809jls1401_4

Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology, 107,* 280–290. doi:10.1016/j.jecp.2010.04.017

Guo, J., Pang, M. F., Yang, L., & Ding, Y. (2012). Learning from comparing multiple examples: On the dilemma of "similar" or "different". *Educational Psychological Review, 24,* 251–269. doi:10.1007/s10648-012-9192-0

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development, 35,* 193–217. doi:10.1159/000277167

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

Kellman, P. J. (2002). Perceptual learning. In R. Gallistel (Ed.), *Stevens' handbook of experimental psychology*: *Vol. 3. Learning, motivation and emotion* (3rd ed., pp. 259–299). New York, NY: Wiley.

Klein, G. A., & Hoffman, R. R. (1993). Seeing the invisible: Perceptual-cognitive aspects of expertise. In M. Rabinowitz (Ed.), *Cognitive science foundations of instruction* (pp. 203–226). Hillsdale, NJ: Erlbaum.

Kok, E. M., de Bruin, A. B. H., Robben, S. G. F., & van Merriënboer, J. J. G. (2013). Learning radiological appearances of diseases: Does comparison help? *Learning and Instruction, 23,* 90–97. doi:10.1016/j.learninstruc.2012.07.004

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

Krupinski, E. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics, 72,* 1205–1217. doi:10.3758/APP.72.5.1205

Kundel, H. L., & La Follette, P. S., Jr. (1972). Visual search patterns and experience with radiological images. *Radiology, 103,* 523–528.

Kurtz, K. J. (1996). Category-based similarity. In G. W. Cottrell (Ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 790). Hillsdale, NJ: Erlbaum.

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1303–1310.

Kurtz, K. J., & Dietrich, E. (2013). *Construing categories*. Manuscript submitted for publication.

Lesgold, A. M. (1984). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory* (pp. 31–60). San Francisco, CA: Freeman.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 732–753. doi:10.1037/0278-7393.24.3.732

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review, 6,* 586–597. doi:10.3758/BF03212967

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565–572.

Lovett, A., Gentner, D., Forbus, K., & Sagi, E. (2009). Using analogical mapping to simulate time-course phenomena in perceptual similarity. *Cognitive Systems Research, 10,* 216–228. doi:10.1016/j.cogsys.2008.03.006

Lowe, R. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education, 14,* 225–244. doi:10.1007/BF03172967

Markman, A. B., & Gentner, D. (1993a). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32,* 517–535. doi:10.1006/jmla.1993.1027

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

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24,* 235–249. doi:10.3758/BF03200884

Markman, A. B., & Gentner, D. (2000). Structure-mapping in the comparison process. *American Journal of Psychology, 113,* 501–538. doi:10.2307/1423470

Miller, J. (1988). Components of the location probability effect in visual search tasks. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 453–471. doi:10.1037/0096-1523.14.3.453

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316. doi:10.1037/0033-295X.92.3.289

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131,* 5–15. doi:10.1037/0096-3445.131.1.5

Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science, 36,* 1019–1050. doi:10.1111/j.1551-6709.2012.01250.x

Samei, E., Flynn, M. J., Peterson, E., & Eyler, W. R. (2003). Subtle lung nodules: Influence of local anatomic variations on detection. *Radiology, 228,* 76–84. doi:10.1148/radiol.2273020509

Schyns, P. G., Goldstone, R. L., & Thibaut, J. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1–17. doi:10.1017/S0140525X98000107

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 31, pp. 305–349). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(08)60413-2

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 681–696. doi:10.1037/0278-7393.23.3.681

Shaw, M. L. (1978). A capacity allocation model for reaction time. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 586–598. doi:10.1037/0096-1523.4.4.586

Shaw, M. L., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 201–211. doi:10.1037/0096-1523.3.2.201

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51,* 599–606. doi:10.3758/BF03211656

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81,* 1768–1786. doi:10.1111/j.1467-8624.2010.01509.x

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136. doi:10.1016/0010-0285(80)90005-5

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. doi:10.1037/0033-295X.84.4.327

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics, 56,* 495–500. doi:10.3758/BF03206946

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–281. doi:10.1207/s15516709cog1802_2

Wisniewski, E. J., & Middleton, E. L. (2002). Of bucket bowls and coffee cup bowls: Spatial alignment in conceptual combination. *Journal of Memory and Language, 46,* 1–23. doi:10.1006/jmla.2001.2801

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238. doi:10.3758/BF03200774

Wood, B. P. (1999). Visual expertise. *Radiology, 211,* 1–3.

Yantis, S. (1996). Attentional capture in vision. In A. Kramer, M. Coles, & G. Logan (Eds.), *Converging operations in the study of selective visual attention* (pp. 45–76). Washington, DC: American Psychological Association. doi:10.1037/10187-002