1 Introduction

Facial expressions of emotion (FEEs) have evolutionary significance because they signal an individual’s emotional state and/or intentions, thereby facilitating social communication. Although many different FEEs exist, there is pan-cultural agreement on the facial portrayal of six FEEs: anger, disgust, fear, happiness, sadness, and surprise. These six ‘basic’, or ‘primary’, FEEs are visually recognized by sighted, neurologically intact individuals belonging to different cultures and races around the world (Ekman and Friesen 1971; Ekman et al 1969).

Ekman and Friesen (1975) proposed that each basic emotion is associated with a distinct musculoskeletal configuration, which is more or less invariant across faces. For example, surprise is commonly expressed by a dropped jaw, raised eyebrows, wide-open eyelids, and a horizontally wrinkled forehead. Sadness is normally portrayed by a protruding lower lip, a downturned mouth, and a slight furrowing of the forehead. The recognition of these emotion-specific facial patterns helps determine the emotional state of an individual, even when looking at an unfamiliar face. This sensitivity to invariants in FEEs is believed to develop early in sighted neurologically intact individuals, with some basic FEEs being recognized in individuals who are seven months of age (Nelson et al 1979).

Although FEEs are normally processed by the visual system, the possibility that blind individuals can decode FEEs has been raised by a growing body of work in haptics. This line of research indicates that manual face exploration is a viable sensory channel for perceiving facial expressions of emotion. For example, in a study...
by Lederman et al (2007), young sighted adults were blindfolded and asked to place their hands on the face of a live actor who had been trained to perform Ekman’s six primary emotions on demand. The actor performed each expression of emotion both statically (the FEE was maintained without change while the participant explored the actor’s face for up to 12 s), and dynamically (the actor performed up to four neutral-to-targeted expression cycles). Performance accuracy was 52% and 74% for the static and dynamic displays, respectively (chance = 17%). Haptic performance also varied according to facial expression of emotion: happiness, surprise, and sadness were recognized at much higher accuracy levels than fear, disgust, and anger.

Live faces are excellent haptic stimuli because they contain a number of cues that are accessible to the hands, as well as to the eyes, including 3-D musculoskeletal structure, skin compliance, and the geometric shapes formed by the features (Kilgour and Lederman 2002). However, touching a peer’s face is a highly intimate act and may neither be appropriate nor possible. Fortunately, surrogate face-like displays such as 3-D facemasks and 2-D raised-line drawings offer potential utility for haptic face processing. In a study by Baron (2008), basic FEEs portrayed on life-like 3-D facemasks were haptically classified by manual exploration. An impressive 82% accuracy was documented for upright displays, despite the absence of some of the cues that are readily available in live faces. In addition, Lederman et al (2008) demonstrated that with little training, the six basic FEEs can even be identified by haptically exploring very simple 2-D raised-line drawings composed solely of contours depicting the external outline of the face, eyebrows, eyes, nose, and mouth. FEEs presented in an upright orientation were classified with accuracies close to 60%, where chance was 14%. Raised-line displays are challenging haptic stimuli because they offer only limited haptically accessible cues; moreover, the spatially distributed contours must be extracted sequentially, thus posing a considerable strain on spatio-temporal integration and memory. On the other hand, tangible graphics, such as 2-D raised-line drawings, are a viable source of useful information for the visually impaired (Heller 1989; Heller et al 2005; Lederman and Campbell 1982). This possibility motivated the present research.

The present study explicitly addresses the effect of training on the haptic classification of FEEs in 2-D raised-line drawings. In addition to assessing the magnitude of learning, we investigated the nature of the underlying processes. More specifically, we considered whether learning involves: (i) recognizing only specific configurations of an actor and an emotion (ie ‘token-specific learning’), (ii) recognizing regularities in each category of facial expression (ie ‘type learning’ or ‘generalization’), or (iii) some combination of token-specific learning and generalization. To the extent that generalization also occurs, we may further conclude that participants learn to perceive invariants associated with a specific FEE category, and that they can successfully transfer this knowledge for purposes of classifying that same emotion expressed by an unfamiliar face.

We present two experiments. In experiment 1, we assessed a group of sixty sighted blindfolded participants. From previous work, we know that, with only minimal training, inexperienced sighted blindfolded individuals can haptically classify FEEs in these displays with accuracy levels well above chance (Lederman et al 2008). Moreover, McGregor et al (2010) recently reported that training can improve the ability to haptically identify individual faces portrayed in 2-D raised-line drawings. Therefore, we expected that some learning would occur if training were provided in a haptic FEE classification task involving the same type of simple 2-D representations. However, both the level attainable with training and the generality of that training are open issues, and thus constituted the primary focus of the current study. The 2-D facial drawings are particularly useful for assessing learning because with such stimuli, initial performance, while above chance, is far from error-free, thus creating a range for improvement to occur.
In experiment 2, we tentatively assessed the ability of adventitiously blind individuals to haptically decode FEEs. Our findings are reported in three case studies.

A hybrid paradigm involving both pre-training tests and post-training tests with old and new (untrained) test displays was used in the experimental design of both experiments. Participants were first tested on their ability to haptically classify six basic FEEs with no prior training. Next they were extensively trained by the experimenter. Finally, they were re-tested, both on their ability to classify the drawings on which they were trained, and on completely novel drawings that portrayed the same six expressions.

Discussion of the results for both experiments is deferred to section 4.

2 Experiment 1. Haptic classification of FEEs by sighted blindfolded participants

2.1 Method

2.1.1 Participants. There was a total of sixty participants (seventeen males, forty-three females, mean age = 19.75 years, SD = 3.43 years). Fifty-five were undergraduate students at Queen’s University, four were graduate students, and one was a professor. All reported normal (or corrected-to-normal) vision, hearing, and sensorimotor hand function, and no injuries to their hands or arms. Fifty-four participants were right-handed, five left-handed, and one ambidextrous. Handedness was assessed with the Edinburgh Handedness Inventory (Oldfield 1971).

2.1.2 Materials and apparatus. Four female actors(1) (ages 20, 22, 29, 69 years) were trained to produce six basic FEEs (anger, disgust, fear, happiness, sadness, and surprise) based on the intense versions of the expressions identified by Ekman and Friesen (1975), plus a neutral expression. The actors’ expressions were recorded with a digital camera. Next, the photographs were modified with Adobe Illustrator v10. Using the pencil tool, we traced primary facial features (eyebrows, eyes, nose, mouth, and external facial outline) with a ‘2 pt’ line width. The photograph of the actor’s face was then deleted so that only the traced facial features remained (figure 1). 2-D raised-line versions of the 28 faces were created by transferring the contour images onto 21 cm × 30 cm swell paper. This type of paper contains particles that burst when exposed to heat, and creates raised lines ~0.5 mm high and ~0.3 mm wide. The size of the actual faces ranged from as small as 15 cm × 13.5 cm to as large as 20 cm × 13.7 cm (mean height = 17.1 cm, mean width = 13.5 cm).

The 4 neutral faces were used to familiarize participants with the task, but were not included in the formal study. The remaining 24 actor × emotion configurations (4 actors × 6 FEEs) were used to create two pairs of mutually exclusive stimulus sets. Each pair of sets was created by quasi-randomly selecting two (of the four possible) configurations for each of the six FEEs, and then adjusting the selections so that each actor appeared an equal number of times per set. In this way, each set contained 12 configurations, with each emotion represented twice and each actor represented 3 times. Set assignment was counterbalanced across participants (fifteen participants per set). The use of 2 pairs of stimulus sets in the formal experiment served to minimize the potential effect of set-specific idiosyncrasies.(2)

(1) Two of the four actors were trained stage performers. We are confident that actor experience did not influence our results for the following reasons: (i) all four actors were given extensive training, (ii) the more experienced actors helped train the inexperienced actors, (iii) all four actors were told to base their FEEs on exaggerated versions of FEEs identified by Ekman and Friesen (1975), and (iv) the coarse nature of the stimuli (static outlines of facial features) made it difficult to capture subtle differences in the expressions of more experienced actors. We further reduced the risk of actor-specific results by including each actor in each stimulus set an equal number of times.

(2) The effect of set assignment was considered in initial statistical analyses, but as no effects were found, data were pooled over that factor because it is of no theoretical interest.
A clipboard taped securely to a table was used to stabilize the raised-line drawings while these were being manually explored. Participants wore a blindfold and headphones with low-volume static noise during trials to block out visual and auditory cues. Participant responses and response times (in ms) were recorded with a custom-designed computer program. A stopwatch was used to limit trial duration to the maximum response time of 45 s. Finally, a brief questionnaire was administered upon completion of the study.

### 2.1.3 Experimental design

A hybrid paradigm involving both pre-training tests and post-training tests with old and new (untrained) test displays was used in the experimental design (refer to figure 2 throughout this section). The first stage of the learning paradigm was pre-training (12 trials). This introduced a basic set of 12 depictions of FEEs (2 actors per FEE), denoted the ‘old set’. The purpose of this stage was to determine how well participants could haptically identify FEEs portrayed in 2-D raised-line drawings prior to training.

Four sub-stages of training with the old set followed in sequence: label-1 (12 trials), feedback-1 (24 trials), label-2 (12 trials), and feedback-2 (12 trials), for a total of 60 training trials. In the label sub-stages, the experimenter named the emotion for the participant before the start of each trial; no response was required. During the feedback sub-stages, participants were required to report the name of the FEE, following which the experimenter gave the correct name of the emotion. Only feedback-2 data were used in the final analysis, denoted ‘final-training’, to maximize learning.

The post-training stage of the learning paradigm, with 24 trials, followed the 60 training trials. In this stage, participants were tested on their ability to classify each of a series of FEEs, 12 belonging to the set on which they were trained (old set).
and 12 novel, comprising 2 portrayals of each of the FEEs by an actor other than the one that had been trained (new set). The order of presentation of old and new items was randomized. Performance on old items, denoted post-training (old), assessed the participant's token-specific learning, whereas performance with new items, denoted post-training (new), assessed generalization (i.e., type learning). Because the stimuli belonged to closed sets of 12 (2 versions of each motion), we were concerned that if participants were presented with a stimulus set only once during each sub-stage of training, they might begin to make classification judgments based on a process of elimination, in turn reducing the value of chance for that stage or sub-stage. To reduce the risk of this occurring, we deliberately varied the number of trials in the two training sub-stages in which accuracy was recorded (24 trials in feedback-1 and 12 trials in feedback-2). The goal was to make it more difficult for participants to predict the end of a sub-stage and to use a counting strategy. For similar reasons, the 12 old set and 12 new set tokens were randomly intermixed and presented in the post-training stage.

2.1.4 Procedure. Written instructions were provided before the start of the study and also before beginning new trial types; specifically, before the start of pre-training and before each of the first two sub-stages of training (label-1 and feedback-1). Verbal instructions were given before the start of label-2, feedback-2, and post-training. A preliminary familiarization stage preceded the formal study. Prior to the start of this stage, participants were told that they would be presented with a drawing depicting a ‘neutral’ face, i.e., devoid of any emotional expression. Participants were told to explore the drawing for as long as they wished, and to remove their hands and say “neutral” when they felt ready to begin the formal experiment. Each trial of the formal study began with the experimenter placing an upright drawing on the table in front of the participant and saying “start”, at which point the participant was free to start exploring it. The specific procedure varies with the stage, as follows. During trials of the pre-training stage, participants were asked to identify the facial expressions portrayed by the drawings as quickly and accurately as possible and were given no feedback following their response. The trial ended when the subject responded with the FEE.
or after 45 s. During the label sub-stages of the training stage, the experimenter labeled
the expression portrayed in each drawing before allowing the participant to explore it.
The participant was instructed to learn what each expression felt like, and the trial
ended when the participant removed his or her hands from the drawing or after 45 s.
During the feedback sub-stages of the training stage, as during the pre-training stage,
participants were asked to identify the expressions as quickly and accurately as possible.
However, the experimenter now provided feedback by naming the expression after the
participant’s response. Trials of the post-training stage were identical to those of the pre-
training stage.

Response time was recorded for all four learning components and was measured
from the time when the experimenter said “start” to when the participant responded. If
the 45 s trial duration limit was exceeded, an RT of 45 s was recorded.

2.2 Results
The $z$ for ANOVAs and for the LSD pairwise tests was set at 0.05. The Greenhouse –
Geisser correction procedure was used when Mauchly’s test of sphericity was violated.
All significant effects and interactions were analyzed a posteriori with LSD pairwise
tests. One-tailed $t$-tests were used to make comparisons against chance. Unless meaningful
to the discussion, non-significant results are not reported.

2.2.1 Accuracy. The mean accuracy (proportion correct) for each emotion, by learning
component, was calculated for each participant. A two-way, within-subjects repeated-
measures ANOVA with learning component (4 levels) and emotion (6 levels) as
within-subjects factors was performed. The main effect of learning component was
highly significant ($F_{3,177} = 88.44$, $p < 0.001$, $\eta_p^2 = 0.60$). As can be seen in figure 3,
accuracy was significantly lower for pre-training than for either final-training or post-
training (old) ($p < 0.001$), or for post-training (new) ($p < 0.01$). This represents an
improvement of 81% from pre-training to final-training, 76% from pre-training to post-
training (old), and 20% from pre-training to post-training (new). Also, performance was
significantly lower for post-training (new) than for either final-training or post-training
(old) ($p < 0.001$).

![Figure 3. Mean accuracy (proportion correct) in experiment 1 by learning component, for each
of the six FEEs. The error bars show ±1 SEM.](image)

The main effect of emotion was also highly significant ($F_{5,295} = 70.65$, $p < 0.001$,
$\eta_p^2 = 0.55$). The order of preference for FEEs from best to worse was: surprise and
happiness, followed by anger and sadness, and finally, disgust and fear (figure 3).

Finally, the learning component × emotion interaction was significant ($F_{15,885} = 8.18$,
$p < 0.001$), although the effect size was relatively small ($\eta_p^2 = 0.12$). All six emotions
tended to follow a common pattern across the four learning components: performance
was lowest at pre-training, higher at final-training and post-training (old), and descended again at post-training (new) (figure 3). The interaction reflects two general trends. First, learning was greater for FEEs that had lower pre-training performance. Accuracy for surprise and happiness was already very high at pre-training (81% and 77%, respectively), reducing the potential for learning, whereas t-tests revealed that pre-training accuracy for anger, disgust, and fear was no greater than chance (all $p$s $> 0.05$), contributing to greater improvement. Thus the magnitude of the training effect varied across emotions. Further, the interaction reflects the fact that generalization to new items was observed, but not consistently across emotions. Anger and sadness were recognized significantly more accurately in post-training (new) than in pre-training ($p$s $< 0.001$, 0.01, respectively), constituting generalization. For the remaining emotions, however, performance on new items essentially reverted to pre-training levels (all $p$s $> 0.40$). Moreover, the performance on new items was below that on old items consistently across emotions (mean difference $= 0.23$), even for those that demonstrated generalization. Thus, learning as measured by accuracy was not broadly generalized.

A subsequent error analysis of stimulus-response confusion matrices indicated that, generally speaking, there were more confusions for disgust, fear, and sadness than for the other emotions, and that these expressions were often confused with each other. Anger was highly confused with surprise in pre-training, but not in subsequent stages of the study. It is also noticeable that, with the exception of anger in pre-training, no strong response biases were observed.

### 2.2.2 Response time.

Out of the total 2880 trials that comprised the four learning components [pre-training, final-training, post-training (old), post-training (new)], 137 were stopped by the experimenter at 45 s (ie at the maximum response time), or $\sim 5\%$. Given the low percentage of stopped trials, the RTs for these trials were included in the analysis with values of 45 s. The same learning component $\times$ emotion two-way ANOVA was performed, this time with response time as the dependent variable. There was a highly significant effect of learning component ($F_{1,177} = 47.28$, $p < 0.001$, $\eta^2_p = 0.45$).

As can be seen in figure 4, participants took significantly longer to respond in pre-training than in the three subsequent learning components (all $p$s $< 0.001$). This represents a 28% decrease in RT from pre-training to final-training, 30% from pre-training to post-training (old), and 20% from pre-training to post-training (new). Also, response time for post-training (new) was longer than for either final-training or post-training (old), which were statistically equal in duration.

![Figure 4](image.png)

**Figure 4.** Mean response time in experiment 1 by learning component, for each of the six FEEs. The error bars show $\pm 1$ SEM.
The main effect of emotion was also highly significant ($F_{4,11295} = 29.28, p < 0.001,$ $\eta^2_p = 0.33$) (figure 4). Surprise was classified faster than any of the other emotions (all $ps < 0.001$, except happiness, $p = 0.01$). Response time for happiness was faster than for all other emotions (all $ps < 0.001$), with the exception of surprise. Anger was recognized faster than disgust and fear ($ps < 0.05, 0.01$, respectively), which were equivalent. Finally, sadness was classified faster than fear ($p < 0.05$).

In contrast to the accuracy analysis, the interaction term in the RT ANOVA was not significant. Thus, the learning trends in response time, including improvement in new tokens relative to pre-training, were statistically equivalent across all six emotions (figure 4). This suggests that generalization may apply more broadly than reflected by accuracy alone.

3 Experiment 2. Haptic classification of FEEs by four blind individuals
3.1 Method
3.1.1 Participants. Three adventitiously blind individuals were recruited through personal contacts. Table 1 provides their demographic information, as well as the history and nature of their visual loss. All three were right-handed and read Braille with their dominant hand at a Grade 2 level, except KB, who read at a Grade 1 level.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age/years</th>
<th>Gender</th>
<th>Highest level of education attained</th>
<th>Visual capacity</th>
<th>History of visual loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH</td>
<td>19</td>
<td>F</td>
<td>undergraduate</td>
<td>some light perception in both eyes</td>
<td>gradual loss since birth due to genetic disease</td>
</tr>
<tr>
<td>SC</td>
<td>19</td>
<td>F</td>
<td>undergraduate</td>
<td>none in either eye</td>
<td>cancer at 5 months with gradual loss until 14 years of age</td>
</tr>
<tr>
<td>KB</td>
<td>34</td>
<td>F</td>
<td>postgraduate (law)</td>
<td>none in right eye; 10% in left eye; can’t count fingers at close distance</td>
<td>congenital glaucoma; some vision in right eye until 25 years of age</td>
</tr>
</tbody>
</table>

3.1.2 Materials and apparatus. The materials were identical to those used in experiment 1, involving sighted individuals, with the exception that the blind participants were not blindfolded, at their request. KB, who had some residual light perception, wore a pair of liquid crystal (LCD) glasses (PLATO®), the lenses of which were opaque during the entirety of the experiment proper. RH could also perceive light, but her eyes were not covered because she felt that wearing the glasses would be uncomfortable.

(3) A fourth, congenitally blind participant, SK (complete vision loss within the first three months of life due to retinopathy of prematurity) was also recruited. Her demographic information follows: 61-year-old female, high school diploma, normal hand function. SK’s accuracy and RT are denoted by a grey line in figures 5 and 6. SK performed with considerably lower accuracy relative to the adventitiously blind and sighted groups throughout the study; nevertheless, her pattern of learning was similar to that of other participants, both sighted and late-blind. Her results are not discussed because we are unable to attribute any observed differences in performance to early vision loss based on this one participant.
3.1.3 **Experimental design and procedure.** The experimental design and procedure were almost identical to those used in the haptic study with sighted subjects (experiment 1). However, unlike the sighted group, all three blind participants were presented with the same old and new stimulus set. We also assessed their hand function prior to commencing the formal study. Were they to show poor haptic processing of FEEs, their performance could have been attributed to impaired hand function, as opposed to loss of vision. Tests of hand function included the 2-point touch test for spatial acuity and the Von Frey filament test for force sensitivity, both of which were conducted in the centre of the index finger pad, and used the method of limits (with 5 alternating ascending and descending series, each terminating after 5 changes in direction), and haptic common-object identification (for further procedural details, see Kilgour et al 2004).

3.2 **Results**

3.2.1 **Tests of hand function.** Threshold results for 2-point touch spatial acuity and pressure sensitivity (individual thresholds based on the mean of the values corresponding to 5 changes in each direction) are shown in table 2. SC and KB correctly classified all 17 objects in the haptic object-identification task. RH gave one incorrect answer: she mislabeled a wine cork as a 'game piece'. The mean response time for all three participants was $\leq 2$ s. These results are indicative of normal hand function (see Jones and Lederman 2006).

### Table 2. Mean 2-point touch and pressure sensitivity thresholds for adventitiously blind.

<table>
<thead>
<tr>
<th>Participant</th>
<th>2-point touch/mm</th>
<th>Force sensitivity/g</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>left hand right hand</td>
<td>left hand right hand</td>
</tr>
<tr>
<td>RH</td>
<td>3.04 2.68</td>
<td>0.008 0.008</td>
</tr>
<tr>
<td>SC</td>
<td>2.24 2.72</td>
<td>0.008 0.008</td>
</tr>
<tr>
<td>KB</td>
<td>4.04 3.52</td>
<td>0.008 0.008</td>
</tr>
</tbody>
</table>

3.2.2 **Accuracy.** Formal statistical calculations by either individual or group were inappropriate for several reasons: (i) the small sample size; (ii) the heterogeneous nature of visual impairment in this sample; and (iii) the discrete nature of the accuracy measure, due to the fact that each emotion was presented only twice within each learning component. It is notable, however, that when each participant’s accuracy is plotted as a function of learning component (figure 5), the emerging learning curves resemble the mean learning trend for sighted participants (thick dark line).

![Figure 5](image_url) **Figure 5.** Mean accuracy of individual blind participants in experiment 2, along with the mean of the sighted group in experiment 1, by learning component. The shaded area represents the 95% confidence interval around the sighted group mean.
As with sighted participants, all three adventitiously blind participants showed clear learning effects for old drawings [mean improvement for post-training (old) = 88%]. RH and SC also improved in their ability to classify FEEs in post-training (new), demonstrating generalization. KB's performance differed from that of the other two adventitiously blind participants in that her accuracy in post-training (new) was slower than in pre-training.

3.2.3 Response time. As with accuracy, no statistical tests were performed. Figure 6 demonstrates that all three adventitiously blind participants performed the task notably faster than the sighted subjects (thick black line). Unlike the sighted group, RH and SC were slower in post-training (new) than they were in pre-training.

4 General discussion

The main purpose of this study was to assess the effect of training on the haptic classification of FEEs in 2-D raised-line drawings, and to examine the nature of the underlying learning processes (ie token-specific and/or generalized learning). We provided participants with systematic training and assessed learning effects.

4.1 The effect of training on task performance by sighted blindfolded participants

As expected, sighted blindfolded participants improved significantly over training in their ability to haptically classify 2-D drawings. Relative to pre-training, accuracy improved at post-training (old) by 76%. This overall increase in accuracy indicates that substantial learning occurred with only 1 h of training. Response times reflected a similar learning trend, tending to be inversely related to classification accuracy; ie participants took longest to respond in pre-training, followed in turn by post-training (new), and finally by final-training and post-training (old), which were equivalent.

Accordingly, we further considered whether participants learned to associate particular drawings with the correct emotion label (ie token-specific learning) and/or to recognize geometric commonalities within FEEs of the same emotion category (ie generalized learning). The results suggest that both types of learning occurred, to varying degrees across emotions.

Response times showed a pattern indicative of generalization, such that, for all emotions, performance on new tokens was superior to baseline but inferior to trained tokens (figure 4). On the other hand, generalization was less evident when measured by performance accuracy. The overall difference in mean accuracy between post-training (new) and pre-training was statistically significant, corresponding to a 20% relative improvement.
This finding indicates that some learning was transferred to successfully classify untrained displays. However, the learning component by emotion interaction reflected the fact that significant improvement was restricted to two emotions: anger and sadness (figure 3). For the remaining emotions, performance with new tokens reverted to pre-training levels.

There are several possible explanations for the fact that generalization was not broadly represented across FEEs, as measured by accuracy. First, it is possible that there were problems with the sensitivity of our accuracy measurement. It appears that with a more sensitive measure, such as RT, all six FEEs generalized. Second, the sparse cues in 2-D raised-line drawings may have restricted the potential for learning in general, thus restricting the potential for generalization. Although happiness and surprise were classified very accurately prior to training, and were consistently classified more accurately than the other emotions [except anger in post-training (old)], they showed no improvement with training. It is possible, then, that the information in displays of happiness and surprise was already exhausted during pre-training, thereby limiting the potential for further learning and generalization. A similar explanation may account for why fear and disgust did not generalize. Although learning for these emotions was significant, the improvement was smaller than for anger and sadness, and may not have been sufficient to allow generalization. As previously suggested by McGregor et al (2010), 2-D raised-line drawings may simply be too sparse to provide the haptic system with sufficient information to make higher-order classifications (Rosch et al 1976), such as those required to recognize the invariant features within each FEE category. Finally, it is possible that further training is needed to see a more pronounced shift from token-specific learning to generalization. In future, exploring the effects of more extensive training would seem warranted.

Overall, some FEEs were haptically identified with higher accuracy than others. Accuracy was highest for surprise and happiness, followed by anger and sadness, and then by disgust and fear. Response times tended to mirror this pattern in that surprise and happiness were identified faster than the other emotions, which were equivalent. This classification performance order is consistent with earlier studies by Lederman and associates (2007, static and dynamic live faces; 2008, 2-D raised-line drawings), and by Baron (2008, 3-D rigid facemasks) in that happiness and surprise were identified with high accuracy, fear and disgust showed low recognition rates, and anger and sadness were intermediate. Similar performance patterns have also been reported in vision research: happiness tends to be classified with greater accuracy than fear or disgust (eg Elfenbein and Ambady 2002; Kohler et al 2004).

How might we account for the different levels of classification accuracy, RT, and mean learning across FEEs? First, a highly recognizable FEE may have distinctive features that set it apart from other FEEs. The more haptically distinctive the configuration of facial features, the easier it is for a perceiver without vision to associate an expression with the correct emotion category. Mouth geometry is especially important for haptic face processing. Ho (2006) assessed the relative importance of specific facial features for the classification of 2-D raised-line FEEs and found that deleting the contours in the mouth region resulted in greater classification impairment than when the eye and eyebrow contours were deleted. As previously suggested by Lederman and associates (2007, 2008), surprise and happiness are easier to classify because they contain distinct mouth geometries (U-shape and O-shape, respectively), and are therefore more perceptually accessible to the hand.

A second factor that may affect accuracy and learning is agreement in feature geometry when the same FEE is depicted by different individuals. Haptically perceptible similarities in the facial portrayal of an FEE by different individuals may signal invariants for that emotion category, and this knowledge can subsequently be transferred to classify emotion in novel faces.
Distinctive mouth geometry and congruence in the portrayal of an emotion across actors may explain why anger and sadness were (a) most improved and (b) showed generalization. For example, looking at figure 1, it appears that there was greater cross-actor congruence in the configuration of the mouth in sad faces than there was in portrayals of fear. For sadness, all four examples featured a closed mouth and downturned lips, whereas fearful mouths varied greatly in their geometry. Similarly, angry faces also showed invariance in mouth geometry: they were characterized by an open mouth that differed from the open mouth associated with surprised faces. Prior to training, anger was frequently confused with surprise, but not vice versa. This may be attributable to an a priori expectation that a wide open mouth signals surprise. With training, the distinct geometry of angry mouths was associated with the correct emotion category. This knowledge was then transferred to successfully classify novel examples of this FEE.

In addition to varying with respect to cue distinctiveness and cue congruence within emotion category, some emotions may be more difficult to classify because they pose a greater strain on cognitive processing than others. Vision researchers have demonstrated that faces portraying a negative emotion, such as sadness, demand greater attentional resources than faces portraying a positive emotion, such as happiness (Eastwood et al. 2001; Srivastava and Srinivasan 2010). As previously discussed, haptic identification of FEEs in 2-D raised-line drawings is cognitively demanding because of the restricted nature and availability of the cues in the drawings themselves. Thus, FEEs associated with a relatively lighter cognitive load may be easier to haptically identify than those posing a greater strain on cognitive resources.

Finally, facial emotions may differ in recognizability because some are recognized more universally than others. Elfenbein and Ambady (2002) conducted a meta-analysis to examine visual emotion recognition within and across cultures. They found that happiness was most recognizable cross-culturally, while disgust and fear were least. Further, the in-group advantage observed (ie emotion was expressed and perceived by members of the same cultural group) was greatest for fear and disgust, and lowest for happiness and anger. It is interesting to speculate that discrepancies in the universal recognizability of primary FEEs may be shared by both haptic and visual systems.

Another issue raised by the present studies is how people are able to recognize 2-D depictions of FEEs prior to training, to the extent that they can do so. The initial success rates in this study differed from chance (17%) for happiness (77%), surprise (81%), and sadness (26%). It is clear, then, that at least some FEEs can be recognized haptically without training. Given the lack of familiarity with this haptic task, observers presumably utilized an a priori representation of each facial expression to interpret the incoming signals from the exploring hand.

To the extent that haptically recognizing FEEs draws on stored categorical representations, it is interesting to consider how this knowledge would be acquired. Past visual experience with FEEs is the most likely source, given that haptic exploration of FEEs is undoubtedly rare in everyday life. The effect of visual experience in our task was probably twofold. First, it may have facilitated mental imaging. For example, a priori representations of happiness may be more available and accessible because smiling is extremely common in the visual world. Second, past visual experience may have encouraged participants to translate their haptic perceptions into visual images and further interpreted these via visual channels. This hypothesis is in line with findings of Lederman et al (1990) that when sighted blindfolded participants were asked to haptically identify 2-D raised-line drawings of 22 common objects, accuracy was significantly correlated with both object imageability and the observer’s ability to visually image (measured by the VVIQ test; Marks 1973). Another possibility is that haptic inputs were interpreted on the basis of modality-independent, higher-level knowledge.
derived from previous experience with the features that differentiate these emotions. For example, it is factual knowledge that happiness is indicated by upturned lips. A third possibility is that representations of facial affect useful for haptic recognition are formed directly from the afferent somatosensory signals available during an individual's own FEE production. However, this hypothesis requires a mechanism by which the feedback from making an expression can become associated with features picked up by the hand during exploration of a surface plane.

4.2 The effect of training on task performance by blind participants

The heterogeneity of subject characteristics and the small sample size (three) of the group of blind participants made statistical analysis inappropriate. Instead, we chose to treat the data as individual cases and to focus on trends and differences relative to the large sighted group.

As can be seen in figure 5, all three adventitiously blind participants improved from pre-training to post-training (old). Therefore, blind participants showed a learning trend indicative of token-specific learning that was similar to that of the sighted. Generalization was also observed for RH and SC, who classified completely novel drawings more accurately in post-training (new) than in pre-training. On the other hand, KB's classification accuracy fell below pre-training levels for novel drawings. There were no obvious differences in the way that KB performed this task. It is possible that the unannounced introduction of novel drawings subsequent to an extensive training period with the old set led to increased confusion. This would also explain why both SC and RH were slower in post-training (new) than in pre-training.

Overall, the adventitiously blind proved to be similar in accuracy to the sighted in pre-training, and tended to be as accurate, if not more so, with subsequent training. Thus our results are in line with those of Heller (1989), who found that, when asked to identify raised-line depictions of common objects, the adventitiously blind outperformed sighted individuals. He attributed this finding to the fact that individuals who lost their sight later in life have had prior exposure to pictures and are generally more experienced in the use of their sense of touch than the sighted. Indeed, in a brief questionnaire implemented after the study was completed, blind participants indicated that they had had some, albeit limited, previous experience with tangible displays (eg 2-D raised-line mathematical charts and graphs, or 3-D tangible maps). Their greater tactual skills probably also contribute to their relatively fast response times (figure 6).

In sum, although limited data for this group make it difficult to interpret the results conclusively, it appears that in just 1 h adventitiously blind individuals can learn to haptically identify 2-D raised-line drawings depicting the six culturally universal FEEs quickly and at levels well above chance. As with sighted participants, learning clearly constituted token-specific learning, and to a lesser extent, generalization. Furthermore, our results indicate that at least some blind individuals outperform sighted persons at this task.

5 Conclusions

Sighted and adventitiously blind participants can be effectively trained to haptically classify the six primary facial expressions of emotion in 2-D raised-line drawings. The learning trend for the specific FEE tokens (combination of actor and expression) was similar for both participant groups. Generalization was also observed for both participant groups. RT for the sighted blindfolded group indicated generalization for all six FEEs; accuracy indicated generalization for anger and sadness. We have addressed a number of possible reasons why full generalization to novel FEEs was not observed with the accuracy measure. Two of the three adventitiously blind participants also classified new drawings more accurately; their RTs did not reflect this generalization.
Although our results for the small adventitiously blind sample tested are limited, they do point to promising interventions, in the form of training emotional recognition for adventitiously blind observers. We further note that several studies suggest that congenitally blind individuals have greater difficulty voluntarily expressing facial affect than the sighted (e.g., Castanho and Otta 1999; Galati et al. 2003). This deficit is attributed to their minimal exposure to and lack of feedback about facial expressions of emotion, which normally occurs via the visual modality (Galati et al. 1997). Given the finding that 2-D facial expressions of emotion contain cues that are also accessible to the hand, and further evidence that visual experience is not essential to make sense of this information, displays like the present ones may provide an effective way for blind individuals to learn about this form of social communication. Systematic training using 2-D raised-line drawings could be used to supplement training on other facial displays, including 3-D facemasks and live faces of the user’s intimates. Preliminary findings with a congenitally blind individual (SK; see grey line in figures 5 and 6) reveal learning patterns similar to those of the sighted blindfolded group, as well as the adventitiously blind group. If in the future, we were able to confirm similar learning patterns in a larger number of congenitally blind individuals, allowing formal statistical analyses, then such applications would be further supported.

Acknowledgments. This research was supported by an Ontario Graduate Scholarship (OGS) awarded to AA, and by a grant from the Canadian Institutes for Health Research (CIHR) to SL. We would like to thank all of our participants, especially our blind population. Thanks to Steve Cutway, for helping with participant recruitment. Finally, we are grateful to Cheryl Hamilton for her assistance with data analysis and her general support.

References
Baron M. 2008 Haptic Classification of Facial Expressions of Emotion on 3-D Facemasks and the Absence of a Haptic Face-Inversion Effect unpublished Undergraduate Honours Thesis, Queen’s University, Kingston, Ontario
Heller M A. 1989 “Picture and pattern perception in the sighted and blind: The advantage of the late blind” Perception 18 379 – 389
Heller M A, McCarthy M, Clark A. 2005 “Pattern perception and pictures for the blind” Psicológica 26 161 – 171

Lederman S J, Campbell J I, 1982 “Tangible graphs for the blind” Human Factors 24 85 – 100


Marks D F, 1973 “Visual imagery differences in the recall of pictures” British Journal of Psychology 64 17 – 24


Oldfield R C, 1971 “The assessment and analysis of handedness: The Edinburgh Inventory” Neuropsychologia 9 97 – 113


Conditions of use. This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.