Cognition & Emotion
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pcem20

The within-subjects design in the study of facial expressions

Michelle Yik [a], Sherri C. Widen [b] & James A. Russell [b]

[a] Division of Social Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong

[b] Department of Psychology, Boston College, Chestnut Hill, MA, USA

Version of record first published: 08 Feb 2013.

To cite this article: Michelle Yik , Sherri C. Widen & James A. Russell (2013): The within-subjects design in the study of facial expressions, Cognition & Emotion, DOI:10.1080/02699931.2013.763769

To link to this article: http://dx.doi.org/10.1080/02699931.2013.763769

COGNITION AND EMOTION, 2013
http://dx.doi.org/10.1080/02699931.2013.763769

Routledge
Taylor & Francis Group

The within-subjects design in the study of facial expressions

Michelle Yik[1], Sherri C. Widen[2], and James A. Russell[2]

[1]Division of Social Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong
[2]Department of Psychology, Boston College, Chestnut Hill, MA, USA

The common within-subjects design of studies on the recognition of emotion from facial expressions allows the judgement of one face to be influenced by previous faces, thus introducing the potential for artefacts. The present study ($N = 344$) showed that the canonical "disgust face" was judged as *disgusted*, provided that the preceding set of faces included "anger expressions", but was judged as *angry* when the preceding set of faces excluded anger but instead included persons who looked sad or about to be sick. Chinese observers showed lower recognition of the "disgust face" than did American observers. Chinese observers also showed lower recognition of the "fear face" when responding in Chinese than in English.

Keywords: Facial expression; Emotion; Disgust; Fear; Method artefact.

According to the Universality Thesis, whatever their age, sex, cultural background, spoken language, or circumstance, human beings recognise certain specific emotions from certain facial expressions (Ekman, 2007; Ekman & Friesen, 2003). Indeed, Shariff and Tracy (2011) characterised doing so as easy and automatic. The Universality Thesis is sometimes cited as if it were well established and non-controversial (see Aviezer, Bentin, Dudarev, & Hassin, 2011). The fate of broad families of emotion theories depends on the fate of the Universality Thesis. The Universality Thesis is a pillar of some theories, inconsistent with others. Support for the Universality Thesis comes largely from high "matching scores", that is, a high proportion of observers agreeing that the person with a smile is *happy*, with the crying face is *sad*, with the disgust nose scrunch is *disgusted*, and so on for each predicted face–emotion pair. The question that motivated the current study is this: Are high matching scores due to a universal human ability to extract emotional information from facial expressions? Or due to method factors in the design of the

Correspondence should be addressed to: Michelle Yik, Division of Social Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: Michelle.Yik@ust.hk

Preparation of this article was facilitated by RGC General Research Fund (644508) to MY and by NSF grant BSC-0421702 to JAR.

We thank Luis Acevedo, Loretta Chan, Francis Ho, Dennis Li, Nicole Nelson, Cheryl Tso, Karin Wang, Matthew Williams, and Kevin Zeng for their help with data collection.

studies that found the matching scores? Or some combination of the two?

Russell (1994) observed that high matching scores came largely from studies with a highly similar method. That method included previewing of faces, within-subjects design, forced-choice response format, multiple examples of each type of facial expression, and posed and extreme facial expressions devoid of information about the expresser's context. For each aspect of this method, altering the method alters matching scores. No one of these method factors alone produces the high matching scores. Nonetheless, each factor could push the scores higher by a small amount such that the cumulative effect (or the interactive effect) of the combination of method factors would be largely, although not entirely, responsible for the high matching scores. The magnitude of the matching scores as they change with method is thus a key issue, for the higher and more robust the scores, the more plausible is the Universality Thesis, whereas the lower the scores as method factors are varied, the more plausible become various alternative explanations, of which there are many (Russell, 1994). Proponents of the Universality Thesis have agreed that matching scores should be high and not dependent on method: "If the Universality Thesis is correct, then those expressions that are pancultural should elicit very high recognition rates [matching scores], generally in the 70–90% range . . . even when methodological constraints are relaxed" (Haidt & Keltner, 1999, p. 238).

Our focus here is on one method factor, the within-subjects design, which is known to introduce threats to internal validity (Shadish, Cook, & Campbell, 2002), but is, all the same, ubiquitous in the study of facial expressions, where it remains little studied. With a within-subjects design, all faces but the first are preceded by one or more other faces typically within a short time. Russell (1994) reviewed all studies cited as tests of the Universality Thesis beginning with Ekman, Sorensen, and Friesen's (1969) report and continuing to the time of the review. All used a within-subjects design with one exception; that exception found that matching scores for facial expressions were on average 20% higher from a within-subjects than from a between-subjects design (Russell, unpublished data, cited by Russell, 1994). A recent review (Nelson & Russell, 2013) of all tests of the universality thesis published since Russell's (1994) review up to 2010 located an additional 21 studies. All used a within-subjects design. Concern with the within-subjects design is part of a larger project concerned with methodological artefacts introduced by any aspect of method that requires observers to view multiple facial expressions, such as previewing and practice trials, during the experiment.

Methodological concern with asking observers to view multiple facial expressions is raised by evidence that the emotion seen in a face depends on what other faces are seen by the observer during the experiment. This claim is known as the Relativity Thesis. Support for the Relativity Thesis has largely been limited to demonstrations that the same face is perceived at different places along broad dimensions of intensity, pleasure, or arousal depending on what other faces (termed anchors) are shown just before or simultaneously (Russell & Fehr, 1987; Thayer, 1980a, 1980b). Some evidence also showed that not just dimensions but the specific category of emotion seen in a face is also relative to what other faces are seen (Pochedly, Widen, & Russell, 2012; Russell, 1991; Russell & Fehr, 1987; Yik & Zeng, 2010). Concern with the within-subjects design is also raised by the suggestion that observers might depend on a process of elimination when they match a face to an emotion label (Russell, 1994). For example, when presented with a series of faces, participants who know how to label the happy and sad faces can eliminate the labels *happy* and *sad* when asked to choose an emotion label for a face for which they do not know the emotion. This process of elimination alters the probability of selecting the predicted label, increasing with the proportion of labels that can be eliminated. The focus in the present study, however, was not on the mechanisms at work in the within-subjects design but simply how much matching scores vary with the within-subjects design.

The study reported here focused on the facial expression said by Ekman and Friesen (1976) to signal disgust. (Disgust was chosen as a strong example of a purported basic emotion with a facial signal and a clear evolutionary rationale for that signal.) Within the research programme centred on the concept of basic emotions, emphasis is placed on the discreteness of each emotion and the specificity of the signal provided by the face. For example, Izard (1994) asked us to:

> imagine an adult and juvenile trampling through the savannah when the leading foot of the adult lands adjacent to a stimulus that should be avoided: a foul-smelling, deteriorating carcass. This happens again, but on the second occasion the stimulus is a deadly viper. A display on both occasions that was consistent only with undifferentiated negative arousal would provide the trailing juvenile with little information for learning to discriminate a disgusting encounter from a terrifying and deadly encounter. Surely, the speed and repertoire of behavioral responses in the two situations need to be different to be adaptive. (p. 291)

In the study reported here, each participant labelled one example of the canonical "disgust" expression taken from Ekman and Friesen's (1976) *Pictures of Facial Affect*. Prior to seeing and labelling the "disgust face", each participant saw and labelled a series of eight other "anchor" facial expressions. Each participant was thus presented with a within-subjects design. The series of prior faces was identical for all participants with one exception: Participants were randomly assigned to one of three between-subjects anchor conditions: In the angry anchor condition, three of the preceding eight facial expressions were different examples of an "anger expression". In the sad anchor condition, the three were different examples of a "sad expression". And in the sick anchor condition, the three were different examples of a "sick face". The "sick face" is an expression being explored in our lab in which the person feels sick and is about to vomit. The sick face would be classified as an example of disgust according to Ekman and Friesen's (1978) criteria, and indeed the modal freely chosen label for this expression is *disgusted* (Widen, Pochedly, Pieloch, & Russell, 2012). But, we believe that naming faces by emotions presupposes what remains to be established (although we do so in this article for convenience). The term "sick face" stems from the instructions given to the actors posing the face to look as if they are about to be sick. The hypothesis to be tested is that the likelihood of perceiving the canonical "disgust face" as expressing disgust varies with anchor condition—an effect possible with a within-subjects design, but not with a between-subjects design. The Universality Thesis, in contrast, predicts that the "disgust face" is perceived as expressing disgust no matter what the preceding faces are.

The Universality Thesis also predicts minimal influence of the observers' cultural background and the language they speak. We therefore also examined cultural and language differences in matching scores and in the effects of a within-subjects design. We completed the study with three groups of observers: Americans responding in English, Chinese responding in English, and Chinese responding in Chinese. The Chinese observers were all bilingual and randomly assigned to respond in English or Chinese.

METHOD

Participants

Participants were 344 university students who participated in exchange for course credit. The sample size for Americans, who were students of Boston College and responded in English, was 117 (57 males; $M_{age} = 20$ years). The sample size for Chinese, who were students at the Hong Kong University of Science and Technology, was 227. The Chinese participants were randomly assigned to complete the study in either English ($N = 114$, 59 males; $M_{age} = 21$ years) or Chinese ($N = 113$, 59 males; $M_{age} = 22$ years).

Materials

Seventeen black-and-white still photographs (1 surprise, 2 happy, 2 fear, 3 anger, 3 sad, and 6 disgust expressions) were taken from Ekman and Friesen's (1976) *Pictures of Facial Affect*. Each

Figure 1. *Examples of facial expressions used in the three anchor conditions: anger, sad, and sick. The photos shown in this figure were not the ones actually used in the experiment.* © *2011 Emotion Development Lab.*

expression was posed by a different person. Three photographs of a "sick face" were developed in our lab (Widen et al., 2012); these photographs too were black and white, and they were not noticeably different in appearance from those of the *Pictures of Facial Affect*. Two of the three showed Action Units 6, 10, and 25 as defined in Ekman and Friesen's (1978) "Facial Action Coding System"; the third showed Action Units 6, 10, and 26. Half the disgust expressions were posed by males, half by females. The preceding anchor faces were chosen so that in each condition approximately half were males, half females. Examples of the types of faces included as anchors are shown in Figure 1.

Dependent measure

The dependent measure was a forced-choice rating scale with seven options: six emotion labels (*happy*, *surprised*, *afraid*, *angry*, *disgusted*, *sad*) plus *none-of-the-above*. For the Chinese version, the emotion words were translated into Chinese with a back-translation procedure. The translations are given in Table 1.

Table 1. *Percentage of participants selecting each option for the "disgust face"*

	Response option						
Sample	*Disgusted e xin*	*Angry fen nu*	*Happy kuai le*	*Sad bei ai*	*Afraid hai pa*	*Surprised jing ya*	*None of the above yi shang mei you yi ge shi he de da an*
The angry anchor condition							
Americans in English	**84.6**	10.3	—	2.6	—	—	2.6
Chinese in English	**63.9**	16.7	—	11.1	—	—	8.3
Chinese in Chinese	**54.1**	24.3	—	2.7	—	—	18.9
The sad anchor condition							
Americans in English	**62.5**	37.5	—	—	—	—	—
Chinese in English	**44.7**	**44.7**	—	2.6	—	—	7.9
Chinese in Chinese	33.3	**56.4**	—	—	—	—	10.3
The sick anchor condition							
Americans in English	34.2	**55.3**	—	—	—	—	10.5
Chinese in English	20.0	**55.0**	2.5	10.0	2.5	—	10.0
Chinese in Chinese	18.9	**62.2**	—	8.1	—	—	10.8

Notes: Americans in English ($n = 117$); Chinese in English ($n = 114$); Chinese in Chinese ($n = 113$). Figures in **bold** indicate the modal response within each sample.

Procedure

Participants were randomly assigned to one of three anchor conditions and, within each condition, to one of six disgust faces, with the proviso of an approximately equal number of participants in each cell. For each participant, there were nine trials. On each trial, participants saw one photograph of a face and rated it on the forced-choice rating scale. That face removed, the next trial began. Every participant received the trials in the same order: X, Happy, Fear, Surprise, X, Fear, Happy, X, and Disgust, with X replaced with one facial expression depending on the anchor condition: angry, sad, or sick.

RESULTS

Our primary data analysis was a four-way analysis of variance (ANOVA): 3 (Anchor Condition: angry, sad, sick) × 3 (Sample: Americans in English, Chinese in English, Chinese in Chinese) × 2 (Sex of Participant: male, female) × 6 (Disgust Face Exemplar). The dependent variable was whether or not the participant labelled the "disgust face" as *disgusted*. Post hoc tests were Scheffé.

Anchor condition

Matching scores for the "disgust face" varied with Anchor Condition, $F(2, 236) = 28.39$, $p < .01$, $\eta_p^2 = .19$. Results are detailed in Table 1 and illustrated in Figure 2. When the preceding set included anger scowls, observers labelled the "disgust face" as *disgusted*. This condition mirrored the typical situation in the within-subjects design and replicated the typical finding. When the preceding set included no anger scowls but included "sad faces" instead, observers labelled the disgust face as *angry* or *disgusted*. When the preceding set included no anger scowls but included "sick faces" instead, even more observers labelled the "disgust face" as *angry*. Scheffé tests showed that matching scores in each anchor condition differed from those in each of the others. Thus, scores in the angry anchor condition ($M = 0.68$, $SD = 0.47$) differed significantly from those in the sad anchor ($M = 0.47$, $SD = 0.50$), which differed significantly from those in the sick anchor condition ($M = 0.24$, $SD = 0.43$). Scores in the angry anchor condition differed significantly from those in the sick anchor condition. There were no interaction effects involving Anchor Condition.

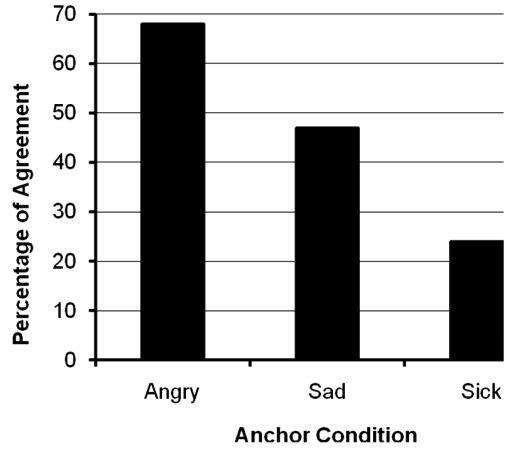

Figure 2. *Matching scores (percentage of agreement on the label disgusted) for the "disgust face" in three anchor conditions.*

Table 1 shows in more detail the distribution of response choices in the three anchor conditions for each sample. For the most part, the "disgust faces" were labelled either as *disgusted* or *angry*, but a few observers selected another alternative. In the angry anchor condition, the modal option was always *disgusted/e xin* in all three samples; in the sick condition, the modal choice was always *angry/fen nu*. In the sad anchor, results were mixed: Americans chose *disgusted/e xin*, Chinese responding in Chinese chose *angry/fen nu*, and Chinese responding in English chose both equally often.

Sample

Matching scores for the "disgust face" varied with Sample, $F(2, 236) = 9.55$, $p < .01$, $\eta_p^2 = .07$. These results are shown in Figure 3. Scheffé tests showed that the Americans responding in English ($M = 0.61$, $SD = 0.49$) scored higher than did the Chinese responding in English ($M = 0.42$, $SD = 0.50$) and the Chinese responding in Chinese ($M = 0.35$, $SD = 0.48$), but the latter two groups

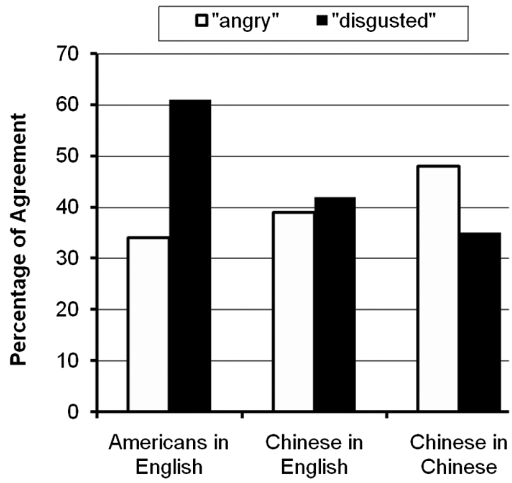

Figure 3. *Percentage of participants selecting disgusted and angry for the "disgust face" in three samples.*

did not differ significantly. Sample did not interact with any other factor in this study

Disgust face exemplar

Some disgust faces were more likely to be labelled *disgusted* than were others, $F(5, 236) = 8.82$, $p < .01$, $\eta_p^2 = .16$. Nonetheless, all disgust faces varied with the conditions of the study in the same manner: Exemplar did not interact with any other factor in this study.

Sex of participant

Female participants were more likely to label the "disgust faces" as *disgusted* than were males, $F(5, 236) = 4.55$, $p < .01$, $\eta_p^2 = .02$. Sex of Participant did not interact with any other factor in the study.

Six other facial expressions

Although not its central focus, this study provided an opportunity to examine matching scores for the faces preceding the target face. Data were analysed with a 2-way ANOVA: 3 (Sample: Americans in English, Chinese in English, Chinese in Chinese) × 2 (Sex of Participant: male, female). The dependent variable was whether or not the participant labelled the face in the conventional manner—*happy* for the "happy face", *afraid* for the "fear face", *surprised* for the "surprise face", *angry* for the "anger face", *sad* for the "sad face", and *disgusted* for the "sick face". Post hoc tests were Scheffé. Results are shown in Table 2. For the "fear face", there was a main effect for Sample, $F(2, 338) = 35.09$, $p < .00$, $\eta_p^2 = .17$. Scheffé tests showed that participants responding in English

Table 2. *Matching scores for six facial expressions by sample*

	Facial expression											
	"Happy"		*"Fear"*		*"Surprised"*		*"Sad"*		*"Angry"*		*"Sick"*	
Sample	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*
Americans in English	0.99 (0.07)	117	0.75[a] (0.33)	117	0.97 (0.18)	117	0.72 (0.19)	40	0.43 (0.24)	39	0.92[c] (0.14)	38
Chinese in English	0.99 (0.07)	114	0.82[a] (0.27)	114	0.96 (0.21)	114	0.74 (0.23)	38	0.55 (0.25)	36	0.73[d] (0.33)	40
Chinese in Chinese	0.97 (0.11)	113	0.47[b] (0.38)	113	0.97 (0.16)	113	0.71 (0.27)	39	0.53 (0.20)	37	0.72[d] (0.29)	37
F-test for sample	$F(2, 338) = 1.70$ $p = .18$ $\eta_p^2 = .01$		$F(2, 338) = 35.09$ $p = .00$ $\eta_p^2 = .17$		$F(2, 338) = 0.20$ $p = .82$ $\eta_p^2 = .00$		$F(2, 111) = 0.17$ $p = .84$ $\eta_p^2 = .00$		$F(2, 106) = 2.93$ $p = .06$ $\eta_p^2 = .05$		$F(2, 109) = 7.60$ $p = .00$ $\eta_p^2 = .12$	

Notes: Matching scores range from 0.00 to 1.00. The data for "happy" and "fear" facial expressions were each based on two photographs; the data for "surprise" expression were based on one photograph; the data for "sad", "anger", and "sick" expressions were each based on three photographs. Different superscripts in one column indicate that the marked pair differs significantly at $p < .01$; same superscripts indicate no significant difference.

scored higher than those responding in Chinese. For the "sick face", there was a main effect for Sample, $F(2, 109) = 7.60$, $p < .00$, $\eta_p^2 = .12$, Sex, $F(2, 109) = 10.59$, $p < .00$, $\eta_p^2 = .09$, and a significant Sample × Sex interaction, $F(2, 109) = 3.77$, $p < .03$, $\eta_p^2 = .06$. Scheffé tests showed that the interaction effect resulted from a significant sex difference among the Chinese responding in English; female participants ($M = 0.88$, $SD = 0.20$) scored higher than male participants ($M = 0.59$, $SD = 0.36$).

In summary, ratings of two of the six facial expressions—fear face and sick face—were significantly different in the different samples. One effect was due to culture and one to language. Culture: Americans were more likely to label the "sick face" as *disgusted* than were Chinese, whether the Chinese responded in English or Chinese. Language: the larger and more intriguing result concerned the "fear face". The two groups responding in English did not differ in labelling the "fear face" as *afraid*, but the Chinese responding in Chinese were much less likely to do so.

DISCUSSION AND CONCLUSION

Our primary concern was methodological. With a within-subjects design, participants view various facial expressions prior to judging all but the first face. (When some faces are previewed or when practice trials are given, then faces are seen prior to all the faces judged.) Here we asked whether the within-subjects design allows the prior set of faces to influence "recognition" of the canonical disgust expression. Within each of three samples (Americans responding in English, Chinese responding in English, and Chinese responding in Chinese), matching scores for the "disgust face" varied significantly with the set of prior faces. Indeed, those scores were highly malleable.

Responding to previous methodological concerns, Rosenberg and Ekman (1995) wrote, "We agree with Russell [1994] that it is important to specify the extent to which methodology can influence judgments of facial emotion. Where we diverge is in the significance we attribute to the changes in agreement levels across methods. We have demonstrated that, with the exception of contempt, variations across methodologies are minor . . . " (p. 134).

The variation in judgements of the "disgust face" seen here was not minor. Not only did the magnitude of matching scores vary greatly, but so did the emotion label chosen by the majority of observers. For each sample, the modal response was *disgusted* in one condition, but *angry* in another. Only one of the six groups listed in Table 1 (Americans in the angry anchor condition) yielded a matching score that passed Haidt and Keltner's (1999) criterion of 70% or better. The other five groups produced matching scores between 18.9% and 63.9%.

The effect seen here of anchor condition was large. But, even if the within-subjects design—or any other one method factor—were found to exert only a minor but reliable effect, all the same, that effect can be important. Russell (1994) argued that no one method factor need be a fatal flaw in studies of facial expressions. Rather, the challenge to validity is the interactive or cumulative impact of various method factors, each exerting pressure in the same direction. Here we held constant various details of method commonly seen in studies of facial expressions, including forced-choice response format, highly selected posed facial expressions, and college student observers. Other evidence indicates that these factors serve to increase matching scores over alternative methods (Naab & Russell, 2007; Russell, 1993a; Wolfgang & Cohen, 1988). Therefore, the impact of the within-subjects design must be understood in the context of other method factors. When the cumulative impact of method factors is considered, the possibility must be faced that the evidence offered in support of the Universality Thesis may depend largely on method.

Our study did not examine the influence of prior faces on labelling a face other than the "disgust expression". The effects seen here join prior evidence to raise but not to answer questions about the use of a within-subjects design on matching scores for other facial expressions

(Pochedly et al., 2012; Russell, 1991; Russell & Fehr, 1987; Thayer, 1980a, 1980b; Yik & Zeng, 2010). We speculate that no face is immune to these effects.

Our study examined only three sets of prior faces rather than all possible sets. One might fault the present findings on the grounds that, in a typical study, different participants see prior faces in different random orders. Perhaps the effects of prior faces therefore counterbalance each other when scores are averaged across participants. There are several problems in this solution. First, not all studies randomise the order of faces for each participant. For example, all of Rosenberg and Ekman's (1995) participants saw the same faces in the same order. Some studies also provide a preview or set of practice trials; again Rosenberg and Ekman's study illustrates this practice. Further, we know of no evidence that randomising or counterbalancing the order does indeed lessen the impact of the within-subjects design. Second, the evidence found in the present study shows that the *types* of faces rather than their order has a major impact on how the target face is interpreted. Third, we have no guarantee that the set of prior faces is the right set—whatever set that might be. If one argues that the right set simply includes all relevant expressions, then no study has met this standard. Many studies have shown participants multiple exemplars of six facial expressions, but is that enough? Here we showed that including the "sick face" in the prior set influenced the results. Therefore, omitting the "sick face" might be said to have influenced all previous results. Should all possible facial expressions be shown first? Perhaps, but there is the practical problem of knowing what is in the set of all possible facial expressions and then showing participants what is likely to be a very large set.

Another limitation of the present study was its focus on method and therefore its inability to show why the within-subjects design influences matching scores. Russell and Fehr's (1987) model of the Relativity Thesis predicted that the sad anchor condition would "push" the disgust face toward higher arousal, the anger anchor condition toward lower arousal. The result would account for increased choice of *angry* in the former condition, and decreased choice in the latter. The Russell and Fehr model would not predict the large differences seen with the sick anchor condition. Another mechanism might be a process of elimination: Once one type of face is assigned to an emotion category, that category might be eliminated for any different but unknown or unclear type of expression. Process of elimination is a plausible account for the lowered use of *disgusted* for the "disgust face" after seeing three "sick faces" and labelling them as *disgusted*. The foregoing speculations await an empirical test.

Whatever may be the mechanism by which the within-subjects design allows prior faces to influence the judgement of a face, we urge caution in the use and interpretation of the within-subjects design. Ekman, O'Sullivan, and Matsumoto (1991a, 1991b) came to the opposite conclusion. Ekman et al. (1991b) wrote, "In our early studies (Ekman, 1972), we found unreliability in initial responses when subjects had to judge expressions . . . . We have found that subjects better understand what is expected of them after trying it a few times. Typically we collect judgments on 30 or more photographs" (p. 294)—in other words, they used a within-subjects design. That observers show "unreliability" in their initial responses means that they failed to select the predicted emotion even with a highly structured response task (selecting one emotion from a short list presented simultaneously with the face); this finding is at odds with the claim that recognition of the emotion signalled by the face is easy and automatic (Shariff & Tracy, 2011). Further, that a within-subjects design allows participants to better understand what the experimenter expects of them is a problem rather than an advantage of the design. In their advocacy of the within-subjects design, Ekman et al. (1991a) reported the results of a study in which 42 participants rated 99 photographs of facial expressions. The participants' task was to select, for each photograph, one emotion term from a list of seven. Ekman et al. (1991a) examined the effect of the one immediately preceding face on the judgement of the "contempt face". The matching scores varied from

52.4% to 76.2%, but Ekman et al. concluded that there was "no effect on the modal judgments of these pictures, regardless of what expression was shown in the immediate anchor condition" (p. 171). These results do not speak to the impact of the full set of preceding faces shown the participants on their judgement of the "contempt face".

Another aspect of method underscored by the present study is the culture of the sample of participants and the language in which the study is conducted. In each anchor condition, the Chinese participants produced lower matching scores for the "disgust face" than did American participants. This finding replicates previous findings from participants in mainland China (Markham & Wang, 1996; Wang & Markham, 1999) and Hong Kong (Yik & Russell, 1999).

The present results can also be thought of as a further demonstration of the effects of context. We distinguish the expresser's context from the context of judgement. When by information about *context* is meant information about the expresser—the situation the expresser confronted, the expresser's body or voice, or the expresser's history—that information concerns factors that plausibly influenced or resulted from the expresser's emotion and are therefore legitimately used by the observer when judging of the expresser's emotion. Even with no expression on the face, someone suffering a devastating insult or preparing to hit another person might well be angry. Thus, observers can legitimately use information about this expresser's context to infer anger. In contrast, when by *context* is meant the methodological context, the context is not a legitimate source of information about the expresser. In the present study, for example, the set of preceding faces provided no genuine information about the expresser.

The present study also has substantive implications. The Relativity Thesis was supported. The emotion seen in the "disgust face" was relative to what other faces had been seen in the experiment. The variation was not simply how intensely disgusted the "disgust face" seemed, but in the discrete category of emotion seen in that face. Relative to a set that included "angry faces", the "disgust face" was judged as *disgusted* by a majority in each sample. Relative to a set that included a "sick face" but not an "angry face", the "disgust face" was judged as *angry* by a majority in each sample. Further, the Relativity Thesis was supported in all three samples, which varied in culture and language.

The Universality Thesis fared less well. Two results here are not what one expects based on the Universality Thesis. First, as just mentioned, the "disgust face" is not a pre-interpreted signal perceived easily and automatically as conveying the same emotion in different methodological contexts. Second, matching scores for the "disgust face" varied with culture. As found in other studies, our Western sample (Americans) fit the prediction of the Universality Thesis better than did a non-Western sample (Chinese). A similar finding also occurred for the "sick face". Serendipitously, we also found an intriguing effect of language. For the "fear face", Chinese responding in English fitted the Universality Thesis better than did Chinese responding in Chinese.

We began by asking whether the high matching scores seen in previous studies are due to a universal human ability to extract emotional information from facial expressions, to method factors in the design of the studies that found these scores, or to some combination of the two. Along with other observers (e.g., Manstead & Fisher, 2002), we remain convinced that the answer is the combination. Although method factors are more powerful than typically acknowledged, they are not the whole story. We do not believe that each facial expression comes with a known meaning in terms of a specific discrete emotion. Thus, the canonical "disgust face" seen in *Pictures of Facial Affect* does not signal disgust, but, all the same, its meaning is not arbitrary. Elsewhere, this position was termed *minimal universality* (Russell, 1995)—meaning that faces universally provide emotion-relevant information, but not a specific emotion. There are various alternative accounts of just what emotion-relevant information a face might provide, such as pleasantness–unpleasantness, social signals, incipient actions, or cognitive appraisals. Method

factors then aid the observer in deciding just what specific emotion to attribute to that face.

Manuscript received 7 August 2012
Revised manuscript received 11 November 2012
Manuscript accepted 30 December 2012
First published online 7 February 2013

REFERENCES

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face–context integration. *Emotion*, *11*, 1406–1414. doi:10.1037/a0023578

Ekman, P. (1972). Universal and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation, 1971* (pp. 207–283). Lincoln, NE: University of Nebraska Press.

Ekman, P. (2007). *Emotions revealed* (2nd ed). New York, NY: Owl Books.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding systems*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face*. Los Altos, CA: Malor Books.

Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991a). Confusions about context in the judgment of facial expression: A reply to "The contempt expression and the relativity thesis". *Motivation and Emotion*, *15*, 169–176. doi:10.1007/BF00995676

Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991b). Contradictions in the study of contempt: What's it all about? Reply to Russell. *Motivation and Emotion*, *15*, 293–296. doi:10.1007/BF00995647

Ekman, P., Sorensen, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*, 86–88. doi:10.1126/science.164.3875.86

Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition and Emotion*, *13*, 225–266. doi:10.1080/026999399379267

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*, 288–299. doi:10.1037/0033-2909.115.2.288

Manstead, A. S. R., & Fisher, A. H. (2002). Beyond the universality-specificity dichotomy. *Cognition and Emotion*, *16*, 1–9. doi:10.1080/0269993014000103

Markham, R., & Wang, L. (1996). Recognition of emotion by Chinese and Australian children. *Journal of Cross-Cultural Psychology*, *27*, 616–643. doi:10.1177/0022022196275008

Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion*, *7*, 736–744. doi:10.1037/1528-3542.7.4.736

Nelson, N., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, *5*, 8–15.

Pochedly, J. T., Widen, S. C., & Russell J. A. (2012). What emotion does the "facial expression of disgust" express? *Emotion*, *12*, 1315–1319. doi:10.1037/a0027998

Rosenberg, E. L., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion*, *19*, 111–138. doi:10.1007/BF02250566

Russell, J. A. (1991). The contempt expression and the relativity thesis. *Motivation and Emotion*, *15*, 149–168. doi:10.1007/BF00995675

Russell, J. A. (1993a). Forced-choice response format in the study of facial expression. *Motivation and Emotion*, *17*, 41–51. doi:10.1007/BF00995206

Russell, J. A. (1993b). *Stimulus presentation in the study of facial expression*. Unpublished manuscript.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *155*, 102–141. doi:10.1037/0033-2909.115.1.102

Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin*, *118*, 379–391. doi:10.1037/0033-2909.118.3.379

Russell, J. A., & Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *Journal of Experimental Psychology: General*, *116*, 223–237. doi:10.1037/0096-3445.116.3.223

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shariff, A. F., & Tracy, J. L. (2011). Emotions expressions: On signals, symbols, and spandrels—A response to Barrett (2011). *Psychological Science*, *20*, 407–408.

Thayer, S. (1980a). The effect of expression sequence and expressor identity on judgments of the intensity

of facial expression. *Journal of Nonverbal Behavior*, *5*, 71–79. doi:10.1007/BF00986510

Thayer, S. (1980b). The effect of facial expression sequence upon judgments of emotion. *Journal of Social Psychology*, *111*, 305–306.

Wang, L., & Markham, R. (1999). The development of a series of photographs of Chinese facial expressions of emotion. *Journal of Cross-Cultural Psychology*, *30*, 397–410. doi:10.1177/0022022199030004001

Widen, S. C., Pochedly, J. T., Pieloch, K., & Russell, J. A. (2012). *Introducing the sick face*. Manuscript in preparation.

Wolfgang, A., & Cohen, M. (1988). Sensitivity of Canadians, Latin Americans, Ethiopians, and Israelis to interracial facial expressions of emotions. *International Journal of Intercultural Relations*, *12*, 139–151. doi:10.1016/0147-1767(88)90045-4

Yik, M., & Russell, J. A. (1999). Interpretation of faces: A cross-cultural study of a prediction from Fridlund's theory. *Cognition and Emotion*, *13*, 93–104. doi:10.1080/026999399379384

Yik, M., & Zeng, K. J. (2010, July). Emotion judgments are relative: Implications for assessing emotional intelligence. In J. Fontaine (Chair), *Recent developments in the assessment of emotional intelligence*. Symposium conducted at the 7th Conference of the International Test Commission, Hong Kong.