A JOURNAL OF THE ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Research Report

Creating Fair Lineups for Suspects With Distinctive Features

Theodora Zarkadi, Kimberley A. Wade, and Neil Stewart

University of Warwick

ABSTRACT—*In their descriptions, eyewitnesses often refer to a culprit's distinctive facial features. However, in a police lineup, selecting the only member with the described distinctive feature is unfair to the suspect and provides the police with little further information. For fair and informative lineups, the distinctive feature should be either replicated across foils or concealed on the target. In the present experiments, replication produced more correct identifications in target-present lineups—without increasing the incorrect identification of foils in target-absent lineups—than did concealment. This pattern, and only this pattern, is predicted by the hybrid-similarity model of recognition.*

Imagine that you witness a crime and the culprit has an obvious marking on his forehead. You would probably feel confident that you could easily identify the culprit from a lineup at a later time. Imagine now that, using your description, the police arrest an innocent man with a similar marking on his forehead. They present you with a photo lineup in which only one person has a marking similar to the one you hold in your memory. Would you identify the innocent suspect as the perpetrator?

Eyewitness research shows that when an innocent suspect matches an eyewitness's description, errant identifications are more likely to occur when the foils do not match the description than when the foils do match the description. Put another way, an innocent suspect who stands out in a lineup is likely to be falsely identified as the culprit (Wells et al., 1998). In simultaneous lineups, in which the individuals are presented all together, eyewitnesses tend to use a relative judgment strategy (Wells, 1984; Wells et al., 1998). In this strategy, the person most closely matching the suspect is selected, even if the overall match is not good. Thus, an innocent suspect with a distinctive feature in common with the culprit is likely to be selected when he or she is the only person in the lineup with that feature. Even if the suspect is actually the culprit, selecting the suspect from a lineup in which only he or she has the distinctive feature reported by the eyewitness offers the police little in the way of new information. After all, the police already know about the distinctive feature from the eyewitness's description, and the eyewitness may be selecting the suspect on the basis of this old information alone.

Identification tests usually consist of a photo array or a video lineup, and police officers typically use one of two techniques to ensure that these lineups are fair and informative. One technique is to replicate the suspect's distinctive feature across lineup members (*replication*), and the other is to conceal the area of the distinctive feature on the face of every lineup member, including the suspect (*concealment*). Both techniques ensure that the suspect does not stand out because of his or her distinctive feature. Although police officers use these procedures daily, and 34% of lineups in England and Wales are digitally manipulated in these ways because the suspects have distinctive features (P. Burton, West Yorkshire Police, personal communication, November 3, 2008), to our knowledge there is no empirical research on the effects of either technique on identification accuracy. Currently, there is no standard regulation giving preference to one technique over the other in the United Kingdom or the United States. Rather, the police officer responsible for each case decides how to construct the lineup that will be presented to eyewitnesses. In Wogalter, Malpass, and McQuiston's (2004) survey of 220 jurisdictions in the United States, 77% of police officers reported replicating distinctive marks across foils, 23% reported adding similar marks to the foils, and 18% said they had tried to conceal the area of the markings. Surprisingly, 30% answered that they did nothing about distinctive features in some cases.

Both replication and concealment make the identification task more difficult for eyewitnesses, as they must rely solely on their memory of other specific facial features. But which technique

Address correspondence to Kimberley Wade, Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom, e-mail: k.a.wade@warwick.ac.uk.

Psychological Science RESEARCH, THEORY, & APPLICATION IN PSYCHOLOGY AND RELATED SCIENCES

allows the police to extract more information from an eyewitness's memory and therefore improve identification performance?

Nosofsky and Zaki's (2003) hybrid-similarity (HS) model of recognition predicts better performance under replication than under concealment. The HS model is a general model of the effects of distinctive features on recognition memory and has been applied to face recognition (Knapp, Nosofsky, & Busey, 2006); thus, it is well suited to modeling these effects. When participants are asked to decide whether they have seen a particular face before, they assess the face's *familiarity*, and this judgment of familiarity determines the probability with which the participants will decide that they have in fact seen the face before. In the HS model, familiarity is defined as the total similarity between the test face and each of the exemplar faces in memory. Similarity between two faces is a joint function of their distance in a large multidimensional space (after Nosofsky, 1986) and their number of shared and unshared discrete features (after Tversky, 1977). Thus, two faces will be similar if they are near one another in the face space, have many discrete features (e.g., scars) in common, and have few unshared discrete features.

Under replication and under concealment, the target face will be, on average, more similar to the exemplars than will a foil. This is because the target matches the exemplar formed when the target was first encountered (hereafter, the *target exemplar*). Therefore, for both techniques, familiarity of the target is higher than familiarity of the foils. However, replication of features across foils at the test exaggerates this difference in familiarity between the target and the foils. Specifically, in the HS model, the common distinctive feature provides a multiplicative boost in the similarity between the target and the target exemplar and also provides a multiplicative boost in the similarity between the foils and the target exemplar. Thus, the absolute difference between the similarities of the target and the foils is increased. Conversely, concealing the target's distinctive features at the test attenuates the difference in familiarity between the target and the foils. So, when the target and foil familiarities are summed with the general familiarity to other faces in memory, the ratio of target familiarity to foil familiarity is higher for replication than for concealment. In summary, replication should increase the difference in familiarity between the target and the foils, whereas concealment should reduce this difference. The HS model, therefore, predicts better performance under replication than under concealment in target-present (TP) lineups. Because common features only boost similarity and missing features only attenuate similarity, the HS model cannot predict the opposite pattern.

In two experiments, we compared replication with concealment. During a study phase, participants viewed a series of faces, a small proportion of which had a distinctive feature. During the test phase, a series of six-person lineups was presented. Experiment 1 used only TP lineups, and participants were forced to select a face. Experiment 2 included target-absent (TA) lineups, and participants were allowed to make a no-identification decision.

STIMULI

The stimuli were developed specifically for this study using photographs of 140 inmates from Florida's Department of Corrections Web site. The selected inmates were 24 years old and had short, brown hair and brown eyes. They were wearing the Department of Corrections uniform and were looking directly toward the camera, exhibiting neutral expressions. The photos showed only inmates' head and neck and were taken against a uniform gray background. None of the inmates wore glasses, and we removed all facial hair, bruises, scars, blemishes, moles, or other identifiers using Adobe Photoshop CS2. We then randomly selected 60 faces and digitally added a distinctive feature to each face using Photoshop. Figure 1 illustrates the six types of distinctive features that we used (a bruise, a tattoo, a piercing, facial hair, a scar, or a mole).

Prior to the experiments, 30 independent judges rated the distinctiveness, attractiveness, and degree of emotional arousal elicited by the 200 faces (80 faces in nondistinctive form only, plus 60 faces in both distinctive and nondistinctive forms). We measured distinctiveness and attractiveness on 9-point Likert scales from 1 (*not at all*) to 9 (*very*). To measure emotional arousal, we used the Self-Assessment Manikin Scale (Bradley & Lang, 1994).

Of the 80 faces that never appeared with distinctive features, we excluded 4 that were outliers on the distinctiveness scale. Of the 60 faces used in both forms, we excluded 6 that were outliers on the distinctiveness scale. There were no outliers on the other scales. We also excluded 2 faces for which there was no difference in distinctiveness before versus after the addition of the distinctive feature.

EXPERIMENT 1

Method

Participants

We recruited 110 students (mean age = 25.5 years, SD = 6.3; 45% female) from the University of Warwick, and they participated voluntarily or received £2 (a little more than U.S. $3). In a within-participants design, participants were presented with both replication and concealment lineups.

Procedure

In the study phase, participants were informed that they would view 32 faces drawn randomly from the stimulus set and would subsequently be tested on their memory of these faces. Participants were asked to view each face carefully. Of the 32 study faces, 6 randomly selected faces had distinctive features (one of each type) and appeared as targets in the test phase. The remaining 26 faces shown during the study phase appeared without distinctive features and were not seen again. The 32 study faces were presented in random order. Each face stimulus was displayed in the center of a computer screen for 2 s.

Fig. 1. Examples of faces used in Experiments 1 and 2 before (top) and after (bottom) the digital addition of a distinctive feature (from left to right: a bruise, a mole, a piercing, a mustache, a scar, and a tattoo).

In the test phase, which followed a 5-min anagram-solving filler task, participants completed a lineup-identification task. They viewed a series of six 6-person lineups and were required to indicate which 1 member of each lineup they had seen in the study phase, indicating their choice by clicking on that member's photo with the computer mouse; they did not have the option of not responding. Participants were instructed that a person previously seen might have a different appearance at test and that their task was to recognize the person previously seen, not the exact photograph. Three of the lineups applied replication (see Fig. 2a), and three applied concealment (see Fig. 2b). The five fillers for each lineup were new, previously unseen faces randomly drawn from the stimulus set. Lineups were displayed in two rows of three photos each (see Fig. 2). The placement of the target in each lineup was determined randomly for each participant, and the six lineups were presented in a random order, which was also determined separately for each participant. There was no time limit for making a decision, and no feedback was provided. The duration of the experiment was approximately 10 min.

Results

Figure 3 shows the proportion of correct and incorrect selections in the two conditions. Participants were significantly more likely to correctly select the suspect when distinctive features

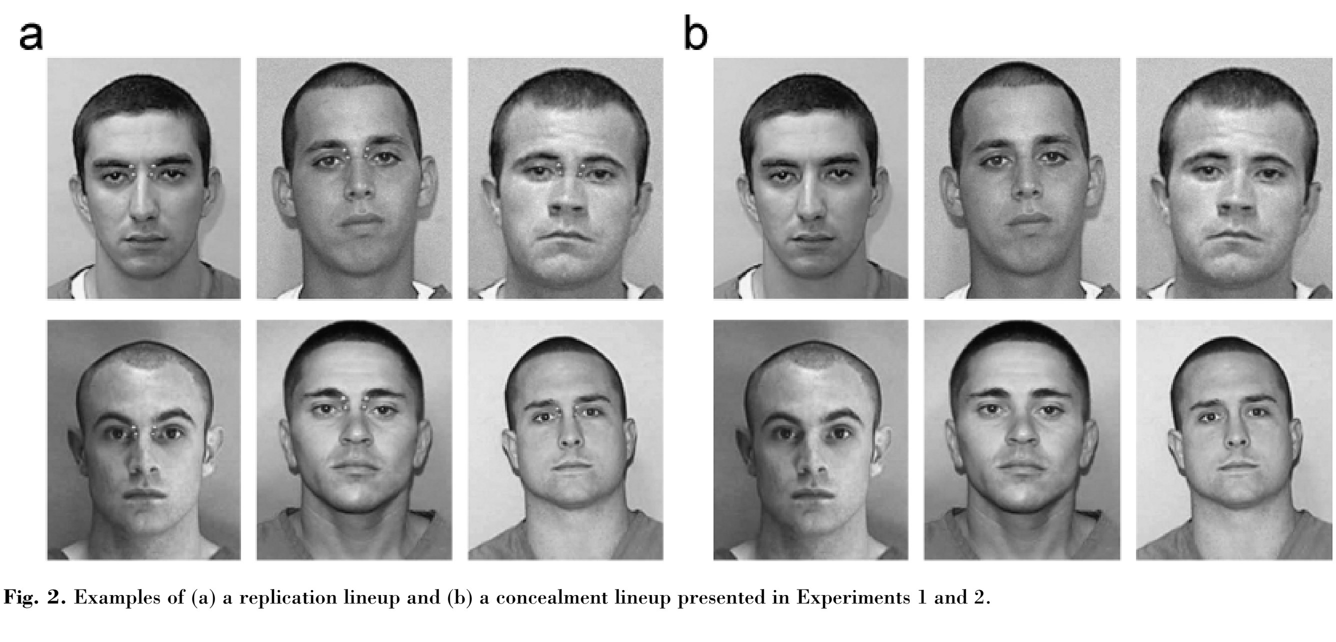

Fig. 2. Examples of (a) a replication lineup and (b) a concealment lineup presented in Experiments 1 and 2.

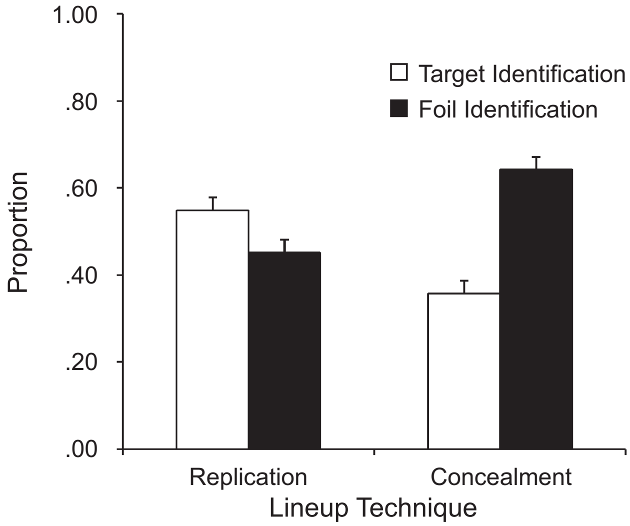

Fig. 3. Mean proportion of correct responses and errors for replication and concealment lineups in Experiment 1. Error bars represent standard errors of the mean.

were replicated across foils rather than concealed on the target, $t(109) = 5.32$, $p < .001$, $p_{\mathrm{rep}} = .99$, $r = .45$.

EXPERIMENT 2

Experiment 2 replicated Experiment 1 and extended the design to include TA lineups. The design was a 2 (lineup technique: replication, concealment) × 2 (target presence: present, absent) within-participants design. In the TA lineups, all six foils were, on average, equally familiar under replication and under concealment (because none of them matched any of the exemplars exactly), so the HS model predicted no difference in identification accuracy between the two conditions for TA lineups.

Method

Participants

A total of 85 psychology students (mean age = 20 years, $SD = 3.0$; 74% female) from the University of Warwick participated for course credit.

Procedure

The procedure was identical to that of Experiment 1, with two modifications. First, in the test phase, participants viewed 12 lineups instead of 6; half were TP and half were TA lineups. Second, if participants recognized none of the faces in the lineup, they were instructed to click on a "none" button below the lineup. TP and TA lineups were randomly intermixed.

Results

Figure 4 shows the proportion of correct and incorrect responses for the replication and concealment techniques. In TP lineups, participants were more accurate at identifying the suspect when distinctive features were replicated across foils rather than concealed, $t(84) = 5.02$, $p < .001$, $p_{\mathrm{rep}} = .99$, $r = .48$; this result replicates the results of Experiment 1. Also, the proportion of errors that were foil identifications (as opposed to nonidentifications) was higher in the replication condition than in the concealment condition, $t(84) = 2.74$, $p < .01$, $p_{\mathrm{rep}} = .97$, $r = .29$. In TA lineups, accuracy did not differ between the concealment and replication conditions: Similar proportions of participants incorrectly selected an innocent foil.

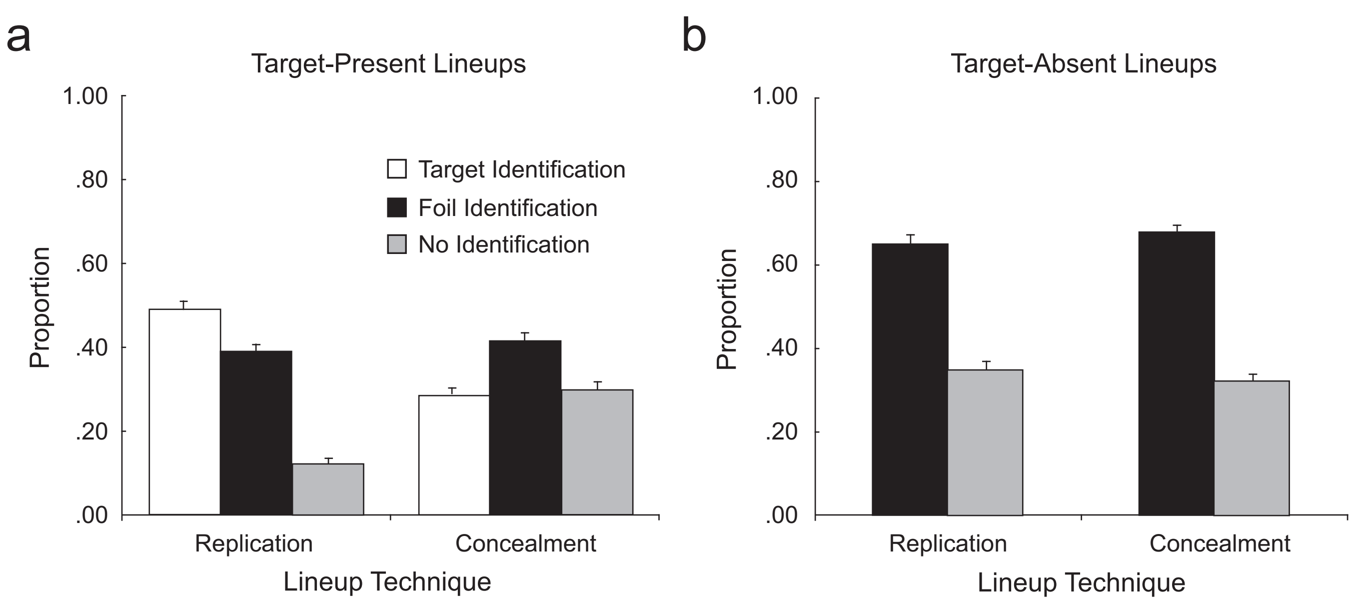

Fig. 4. Mean proportion of correct responses and errors for replication and concealment lineups in Experiment 2: (a) target-present lineups and (b) target-absent lineups. Error bars represent standard errors of the mean.

In summary, our results suggest that replication is better than concealment for constructing lineups because replication increased the probability of selecting the target when a target was present without increasing the probability of selecting an innocent foil when the target was absent. The only drawback was that when the target in TP lineups was not identified, replication (compared with concealment) rendered participants less willing to make a no-identification decision. However, in absolute terms, incorrect foil selections were equally likely for the two techniques.

DISCUSSION

Our finding that correct identifications increased in TP lineups created using replication supports the HS model of recognition memory. Standard global-familiarity models, which do not take into account the effects of distinctive features (e.g., Valentine & Ferrara, 1991), cannot account for our data. Under these models, the target:foil familiarity ratio in TP lineups is the same for concealment and replication lineups. Therefore, standard global-familiarity models predict no difference in identification performance between the two kinds of lineups. This prediction is not supported by our results.

Standard global-familiarity models also predict that participants will make increased false identifications in TA lineups created using the concealment technique: Because faces without distinctive features resemble many other faces without distinctive features seen in the study phase, the overall familiarity evoked should be increased under concealment; hence, participants shown a concealment lineup should have an increased tendency to choose someone from the lineup and to make false identifications. Under replication, the opposite should be true. However, our data revealed no difference in choice rates between replication and concealment lineups that did not include the target.

In Experiment 2, the improvement when distinctive features were replicated rather than concealed came from a reduction in incorrect no-identification decisions. It could be argued that the increase in hits in the replication condition resulted from an increased tendency to select someone from the lineup. Such a mechanism, though, would also generate more false identifications in both TP and TA lineups in the replication condition. However, in both the TP and TA lineups of Experiment 2, participants were as likely to select a person from the lineup in the replication condition as in the concealment condition, despite the fact that targets were correctly identified more often in the replication condition.

Our finding that replication (in which case the suspect remains unchanged between study and test) produces more accurate identifications than concealment (in which case the suspect is altered between study and test) is consistent with the changed-appearance literature. Lineup-identification studies, for instance, show that disguises (Cutler, Penrod, & Martens, 1987a, 1987b; Cutler, Penrod, O'Rourke, & Martens, 1986), changes in hair style or facial hair, and the addition or removal of glasses (Read, 1995) impair identification performance. Likewise, recognition-memory studies show that disguises, changes in pose or facial expression, addition or removal of glasses (Patterson & Baddeley, 1977), changes in visual angle (Bruce, 1982), and the effect of the target's aging (Read, Vokey, & Hammersley, 1990) increase false-identification rates (see also Shapiro & Penrod, 1986).

Our study is directly relevant to cases in which an eyewitness reports a culprit's distinctive feature. Wells and his colleagues argued that when a suspect has a distinctive feature that is not reported, lineups should follow the principle of "propitious heterogeneity" (Luus & Wells, 1991; Wells et al., 1998); that is, the distinctive feature should not be replicated among the foils. However, research suggests that replication should still be applied in such cases. People are able to encode information without concurrent awareness of what is being encoded (Shanks & St. John, 1994). So, although eyewitnesses may not verbalize the presence of a distinctive feature, they may be able to remember it should they see it on the culprit at the time of the lineup. For reasons of fairness, then, everyone in the lineup should have the distinctive feature.

We used a mathematical model of the effect of distinctive features on recognition memory to make predictions for real-world lineups. We predicted that replicating a distinctive feature across foils is better than concealing it on the suspect, because replication amplifies the difference in the familiarity of the target and the foils, whereas concealment attenuates this difference. Two experiments confirmed this prediction. Police officers should be aware of this theoretical and empirical result when constructing lineups for suspects with distinctive features and should replicate rather than conceal these features.

Acknowledgments—We wish to thank Zach Estes and Rob Nash for helpful comments on previous versions of this manuscript. This research was partly supported by Economic and Social Research Council Grants RES-000-23-1372 and RES-062-23-0952.

REFERENCES

Bradley, M.M., & Lang, P.J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*, 49–59.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processing in face recognition. *British Journal of Psychology*, *73*, 105–116.

Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987a). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology*, *72*, 629–637.

Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987b). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior*, *11*, 233–258.

Cutler, B.L., Penrod, S.D., O'Rourke, T.E., & Martens, R.K. (1986). Unconfounding the effect of contextual cues on eyewitness identification accuracy. *Social Behavior*, *2*, 113–134.

Knapp, B.R., Nosofsky, R.M., & Busey, T.A. (2006). Recognizing distinctive faces: A hybrid-similarity exemplar model account. *Memory & Cognition*, *34*, 877–889.

Luus, C.A., & Wells, G.L. (1991). Eyewitness identification and the selection of distractors for lineups. *Law and Human Behavior*, *15*, 43–57.

Nosofsky, R.M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R.M., & Zaki, S.R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1194–1209.

Patterson, K.E., & Baddeley, A.D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 406–417.

Read, J.D. (1995). The availability heuristic in person identification—the sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology*, *9*, 91–121.

Read, J.D., Vokey, J.R., & Hammersley, R. (1990). Changing photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy-confidence relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 870–882.

Shanks, D.R., & St. John, M.F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.

Shapiro, P.N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*, 139–156.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Valentine, T., & Ferrara, A. (1991). Typicality in categorization, recognition and identification: Evidence from face recognition. *British Journal of Psychology*, *82*, 87–102.

Wells, G.L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, *14*, 89–103.

Wells, G.L., Small, M., Penrod, S., Malpass, R.S., Fulero, S.M., & Brimacombe, C.A.E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*, 603–647.

Wogalter, M.S., Malpass, R.S., & McQuiston, D.E. (2004). A national survey of U.S. police on preparation and conduct of identification lineups. *Psychology, Crime, & Law*, *10*, 69–82.

(Received 12/17/08; Revision accepted 3/27/09)