Confidence Judgments on Color Category Best Exemplars

Nancy Alvarado
University of California, San Diego

Kimberly A. Jameson
University of California, Irvine
University of California, San Diego

Although basic color terms and basic color appearances have been shown to produce higher confidence ratings in a variety of naming and judgment tasks, our findings suggest that when different ethnolinguistic cultures are compared, higher confidence is not strictly linked to the basic foci of Berlin and Kay nor the centroid samples identified by Boynton and colleagues. This raises important questions about high confidence as evidence of the salience of basic color foci, a point central to the widely accepted basic color-term theoretical framework. This study analyzes confidence judgment data for Vietnamese and English color naming, suggesting that high confidence may be more directly linked to aspects of a task rather than universal focal color stimuli. Culture-specific patterns of naming, an individual's access to shared cultural knowledge, and goodness of fit between exemplars and names provide a more complete explanation of the higher confidence observed for certain color appearances.

Keywords: *color; color naming; confidence; color categorization; color salience*

Cross-Cultural Research, Vol. 39 No. 2, May 2005 134-158
DOI: 10.1177/1069397104273628

Despite recent controversies, the cross-cultural study of color naming and color categorization is still widely believed to represent a pan-human cognitive universal (see review by Hardin & Maffi, 1997; Kay & Maffi, 1999). Reflecting this view, Guest and Van Laar (2002) commented:

> Much research has determined within and between-culture agreement in what the fundamental names ("basic colour terms" or BCTs) are, and what sensations they denote. . . . Such universals suggest that any results from visual-search experiments involving colour naming could be stable and widely usable. (p. 445)

They further suggested that these basic color sensations share a property of greater "nameability." This nameability or codeability can be measured by indices such as the use of monolexemic terms versus modified terms, the modal frequency and the variability in assignment of names during a naming task (Guest & Van Laar, 2000). In accord with the received view, they asserted that this quality of nameability produces higher confidence judgments, faster response times, and greater accuracy in tasks employing highly nameable color appearances. They present data in English as an initial validation of their construct of universal nameability and supply an explanation that supports the received view of universal focal color salience (Guest & Van Laar, 2000, 2002). In the current article, we use a cross-cultural comparison of naming behaviors in English and Vietnamese to examine the relationship between confidence judgments and nameability and apply our own model of color naming to explain our results.

Although Berlin and Kay's (1969) 11 basic color terms are used more frequently than other color names (with and without modifiers), and some color appearances are unquestionably easier to name than others, contrary to the received view that Guest and Van Laar (2000, 2002) support, our research suggests that universality does not extend to defining a set of specific color appearances that will be most nameable cross-culturally or even across different kinds of tasks. First, our data suggest that the colors considered most nameable in one culture are not necessarily those

---

Authors' Note: *Please direct correspondence concerning this article to Nancy Alvarado, Center for Brain and Cognition, University of California, San Diego, MC-0109, 9500 Gilman Dr., La Jolla, CA 92093-0109; e-mail: nalvarado@csupomona.edu*

considered most nameable in another. Nor are the colors producing highest confidence in naming the same across cultures. Second, the processes involved in making confidence judgments seem to vary cross-culturally and across tasks. This makes confidence an uncertain basis for validating measures of nameability cross-culturally. Third, there appears to be an interaction between the processes of judging confidence and other processes involved in making an accurate speeded response—an issue that complicates the interpretation of response time measures (Baranski & Petrusic, 2001). Correlating response times with accuracy and relating both to confidence or nameability may be misleading because the measures are not independent. Nor is accuracy related to confidence in a straightforward way within or across cultures.

Our data suggest that confidence is a measure of an individual's conformance to a specific culture's normative naming patterns, not strictly a measure of a color exemplar's inherent salience or so-called basicness. We suggest that confidence judgments only indirectly relate to such measures of basic-term nameability as modal term frequency (number of people assigning the same name to a color sample) or variability in naming (the total number of names given to a color sample). Rather, confidence judgments appear to be highest for those stimuli most salient with respect to the dimensions emphasized within the language's color-category-naming system, including, for example, color characteristics named using polylexemic combinations and nonbasic terms. Confidence also appears to be strongly related to the boundaries of the categories lexicalized by a specific naming system. Therefore, we propose that color nameability varies substantially with the naming patterns of an ethnolinguistic culture and is not simply an interaction between the physical properties of a stimulus and the color-processing mechanisms in the human visual system.

We agree that confidence judgments may be a sensitive measure of nameability; however, we suggest that the stimuli producing highest confidence will be different across different ethnolinguistic cultures, as when, for example, a naming system differentiates items based on lightness (brightness) rather than solely emphasizing hue (Alvarado & Jameson, 2002; MacLaury, 1992). Thus, Guest and Van Laar's (2000, 2002) most nameable color stimuli may produce strong results for English, which emphasizes hue, but produce less strong results for Vietnamese, which permits greater use of modified color terms to describe the lightness or brightness dimension (Alvarado & Jameson, 2002). Conversely, a different set of highly

nameable color stimuli may produce stronger confidence results for Vietnamese speakers, but less strong results for English. Because dimensions other than hue may be coded using modifiers or terms typically designated as nonbasic, important cultural differences in nameability may be obscured by the empirical emphasis of the handful of monolexemic basic terms and so-called focal colors that are considered important by established theory.

Our data also show that bilingual individuals fluent in English and Vietnamese tend to apply English naming patterns when making responses in Vietnamese, suggesting that bilinguals represent a different pattern of naming than monolingual individuals speaking either language.

WHAT IS REVEALED BY CONFIDENCE IN A NAMING TASK?

To evaluate and compare confidence judgments about color names across cultures, one must consider what may be taking place cognitively—what cognitive processes form the foundation of the metacognitive judgment. Confidence judgments are usually characterized as a metacognitive evaluation of some internal knowledge state. Confidence is typically assessed in the context of making a judgment under uncertainty. For example, in many experimental contexts, a confidence judgment may estimate the likelihood of making a correct recognition response on a future memory test, or the likelihood of having made a correct judgment on a sensory perception or discrimination task. The request for a confidence rating implies that individuals can and should monitor some internal subjective quality of their own thought that will guide them in estimating the accuracy of their responses. Depending on the nature of the task, this can involve estimating the ease with which an answer was recalled, the sense of familiarity associated with retrieval cues, an estimation of an item's statistical frequency within a stimulus set, the relative difficulty (mental effort) in making a choice between alternatives, and so on. Highly sophisticated models of such processes are being developed for memory and sensory discrimination processes (see, e.g., Juslin & Olsson, 1997; Vickers & Pietsch, 2001); however, none have been described specifically for naming tasks.

The notion of accuracy is inherent to making a confidence judgment and explicitly mentioned in the instructions of many tasks.

Accuracy of response is used to determine how well confidence judgments track actual performance. Guest and Van Laar (2002) determined accuracy by comparing a current response with a previous one. In their tasks, the basis for a participant's confidence assessment varies. Initially it evaluates prospective memory for a choice, then in the response phase of their study, it assesses the ability to guess someone else's previous choice, clearly a higher degree of uncertainty. In their study, confidence judgments do not assess the goodness of naming explicitly, and for this reason their confidence data cannot be considered a direct measure of nameability. Their assumption seems to be that a set of more nameable stimuli will produce higher confidence because greater nameability will enhance performance (reduce errors) on their matching tasks. This is a questionable assumption because the relationship between task difficulty, accuracy, response time, and confidence is known to be complex. Fast response times tend to accompany higher confidence and accuracy; however, longer response times are ambiguous with respect to accuracy and confidence. With more difficult trials, individuals may simply need more time to make a highly accurate and confident response, or they may not produce the correct response despite taking extra time. Furthermore, people can be fast and highly confident while making inaccurate responses, especially when their mistake is normative (e.g., when nearly everyone unknowingly makes the same mistake).

Accuracy is meaningful when it comes to a memory task where correct answers can be objectively determined; however, what constitutes accurate naming when subjective, and most certainly variable, perceptual color experiences are involved? In a naming task, the notion of accuracy implies that there is a right versus wrong name for a given color sample. Because we cannot know the subjective experience of any individual, and because wide color appearance variation is known (e.g., Kuehni, 2004; Webster et al., 2002), assessing accuracy of color naming in any absolute sense is not possible. Nevertheless, individuals are required to make some rating, and they also know that some answers are likely to be considered better than others in naming color appearances. This knowledge encourages them to consider what might pass as a so-called normative answer when making their responses. The characteristics of the stimulus set also provide the participant with implicit information about the fineness of discrimination required by the task and thus the specificity of naming appropriate to the task.

When less discriminable stimuli are presented, more specific names are required to differentiate them verbally. Thus, demand characteristics of the task also establish criteria that shape confidence judgments for the names generated. An individual might evaluate how close her proposed name comes to matching the likely normative or consensual name for a color appearance—as opposed to only estimating how suitable her proposed name is given her own subjective experience of that color appearance. Through these pragmatic features, the notion of accuracy implicitly introduces a standard that is cultural, linguistic, and external to the individual's own experience, and appropriate to the stimulus set and task demands.

IS CONFIDENCE LINKED TO FOCAL OR LANDMARK HUES?

Our research supports the idea that confidence in color naming is based on shared ethnolinguistic knowledge rather than a panhuman visual neurophysiology that privileges certain fundamental color appearances (such as the landmark hues related to Hering's color opponent processing; see Hardin, 2005). In Jameson and Alvarado (2003), we hoped to discover whether the previously noted perceptual salience of certain category focal colors would, in general, contribute to greater agreement in naming under the less constrained naming conditions we used. To examine this issue across language groups, we compared color-naming behaviors in Vietnamese and English, two languages with different linguistic categories for green and blue and yellow and orange. We also investigated the impact of the level of access to terms on naming behavior by including bilingual Vietnamese living in the United States as a third comparison group. We presented participants in these three groups with 110 surface color appearances sampled to systematically represent OSA space (Wyszecki & Stiles, 1982), including a subset matched to Boynton and Olson's (1987) landmark hues and a subset matched to Berlin and Kay's (1969) basic colors (as specified for each language group). These samples were viewed one at a time in random order (viewed under controlled illumination-approximating CIE illuminant C; Wyszecki & Stiles, 1982). We asked participants to name each color without constraint to basic or monolexemic terms, and we asked them to rate their confidence that their assigned name was correct using a scale from 1 to

5 (*least confident* to *most confident*). Bilingual participants responded in Vietnamese. Details of this study are provided by Jameson and Alvarado (2003). We found that when participants were permitted to freely name samples without being constrained to use basic color terms or monolexemes (single words), their naming behavior (a) did not suggest a difference in the salience of focal or landmark colors from that of nonbasic colors and (b) showed group differences in the use of polylexemic names, modifying terms, compounds, and object glosses (described by Alvarado & Jameson, 2002).

In addition, the samples considered most nameable because they produced the highest modal frequency of naming and lower variability in naming (fewer different names assigned to the same sample) did not produce highest confidence. Furthermore, the samples identified as landmark or focal basic hues in previous research (Berlin & Kay, 1969; Boynton & Olson, 1987) produced no higher confidence than other samples, nor were they among the samples producing highest mean confidence.

In sum, our confidence results do not support a theory asserting a strong linkage between focal basic colors and confidence (Guest & Van Laar, 2000, 2002), nor the focal color salience explanation based on pan-human universal color processing that is integral to the received view (Kay & Maffi, 1999).

ARE THE MOST CONFIDENTLY NAMED COLORS THE SAME IN TWO LANGUAGES?

If simple focal color salience is not responsible for color-naming universality, then what explanations can be offered for the relationship between confidence and color naming? In the interest of answering this question, this article presents further analyses of the confidence judgments associated with unconstrained naming in our previous study (Jameson & Alvarado, 2003). Table 1 lists the 10 color appearance stimuli with the highest mean confidence ratings in each language group. First, note that there is little overlap between the highest confidence stimuli for the two monolingual groups but almost complete overlap between the English and bilingual Vietnamese groups (only one color stimulus differs and that was ranked 11th on the Vietnamese list). All of the color stimuli producing the greatest confidence in the monolingual English and bilingual Vietnamese groups tended to be highly saturated in hue.

They were all named using monolexemic basic terms in English and bilingual Vietnamese, except for a green stimlus that was named *xanh la cay* (leaf green) in Vietnamese. The names used with the second highest modal frequency for these stimuli tended to include the modifiers *bright* (in English) or *fresh* (in Vietnamese). In contrast, among the stimuli named with the greatest confidence by the monolingual Vietnamese were four of the highly saturated stimuli found on the other two lists; however, six other color stimuli that varied in lightness (brightness) were confidently named using polylexemic modified and contextualized terms. This color-naming pattern is consistent with the more frequent use of modifiers and compound terms generally found in the Vietnamese language (see Alvarado & Jameson, 2002) but reflects a greater inclusion of the brightness dimension in the patterns of naming that is made possible only by permitting modifier use in naming.

Although responding in Vietnamese, the bilingual participants showed the same tendency as English speakers toward greater confidence when naming highly saturated color exemplars, and less tendency to use compound and modified terms (though more than monolingual English speakers). They present a pattern of naming that combines elements from both languages but tends to more strongly parallel English naming patterns (see also Alvarado & Jameson, 2002 for a description of bilingual modifier use). Although higher than that seen for English speakers, the mean confidence among bilingual speakers compared to both monolingual groups was less strongly correlated with other measures of nameability, including frequency of the modal name and variability of naming, as shown in Table 2. This would not be the case if bilinguals were simply less confident about samples they did not remember the names for. This suggests that the basis for the confidence assessment made by bilingual participants may have been more complicated. Reduced confidence might occur because it was less clear to participants which of the two normative naming criteria to apply, or participants may have been less confident of their command of naming practices in either language (e.g., if they were not using their native language frequently and thus losing competence in it while in the process of acquiring English). As is typically the case in paradigms assessing confidence, different participants may also rely on different criteria to make their judgments. It is interesting to note, confidence judgments for bilinguals were more strongly correlated with the mean frequency and variability measures of English participants than they were with the same

TABLE 1
Comparison of Highest Confidence Exemplars for Each Language Group

English		*Bilingual Vietnamese*		*Monolingual Vietnamese*	
OSA Tile	*Modal Name*	*OSA Tile*	*Modal Name*	*OSA Tile*	*Modal Name*
411	yellow	411[a]	vang (yellow)	186[a]	do (red)
50	blue	106[a]	xanh la cay (leaf green)	411[a]	vang (yellow)
108	green	69	xanh la cay (leaf green)	29[a]	tim (purple)
246	orange	29[a]	tim (purple)	215	xanh da troi (sky blue)
106	green	246[a]	cam (orange)	413[a]	vang tuoi (fresh yellow)
29	purple	413[a]	vang (yellow)	54	tim dam (dark purple)
83	purple	11[a]	tim (purple)	217	tim nhat (light purple)
413	yellow	50[a]	xanh (grue)	340	xanh la cay (leaf green)
186	red	186[a]	do (red)	257	xanh dot choi (banana leaf green)
11	purple	108[a]	xanh la cay (leaf green)	408	vang lot (light yellow)

NOTE: a. Also included among the English highest confidence exemplars (listed in left column).

measures for the bilingual group. In addition, bilingual confidence ratings were less strongly correlated with monolingual Vietnamese frequency and variability than with frequency and variability for the bilingual group.

As described by Jameson and Alvarado (2003), using Spearman's rho, confidence judgments were highly correlated across the English and Vietnamese languages: English versus bilingual Vietnamese, r = .761; English versus monolingual Vietnamese, r = .635; bilingual Vietnamese versus monolingual Vietnamese, r = .654. A regression analysis of confidence on measures of nameability using our English data showed that variability of naming (number of names assigned to a sample) accounted for most of the variance in confidence ratings (R^2 = .648). Modal frequency (number of people giving the same name to a sample) accounted for less (R^2 = .355) and did not add much to the predictive ability of variability alone in a combined model (R^2 = .654). English measures of variability and frequency were better predictors of bilingual Vietnamese confidence than were monolingual Vietnamese measures, supporting our interpretation that the

TABLE 2
Correlations Between Mean Confidence Ratings and Measures of Nameability by Language Group

Language	*Modal Term Frequency*	*Variability*	*Agreement Ratio*
Monolingual English	.60	–.77	.68
Bilingual Vietnamese	.48	–.57	.57
Monolingual Vietnamese	.60	–.71	.66

Spearman's rho, all correlations significant, $p < .001$, two-tailed, $N = 110$.

bilingual participants were using acquired English naming patterns when responding in Vietnamese.

Confidence also varied systematically with color category naming differences in the two languages. This was clearest for the color usually named *orange* in English but called variously *cam* (a gloss for the fruit orange), *vang dam* (dark yellow) or *do* (red) by monolingual Vietnamese speakers. The English category of *yellow* excludes orange and red; however, the Vietnamese category of yellow is broader and may include color appearances called *orange* and *red* in English. The Vietnamese category named using the term *vang* (yellow) is thus more ambiguous than that for English. Bilingual speakers of Vietnamese responding in Vietnamese tended to name *orange* using the term *cam* (orange), as English speakers do. These differences in the linkage between color terms and categories were readily apparent in the confidence ratings of participants naming color samples in the yellow category, as shown in Figures 1 through 3.

Figure 1 selects only those color samples classified as yellow (i.e., named using the term *yellow* or some synonym or composite of yellow) by English speakers. Confidence ratings peak at the OSA lightness (L) level where yellow is highly saturated, consistent with the selection of yellow best exemplars in previous studies (e.g., Berlin & Kay, 1969). Although confidence ratings tend to be higher for Vietnamese speaking participants than for English speakers (see discussion in the next section), the peak in confidence occurs at the same OSA L level in all three ethnolinguistic groups (approximately 30 speakers in each group). As shown in Table 1, all three groups show highest confidence for the same highly saturated exemplar of yellow (OSA tile #411). Figure 2 selects only those color samples classified as *vang* (yellow) by monolingual Vietnamese speakers. Note that the color samples

included with this change in selection criterion span more lightness levels (extending to –3 instead of 0). This occurs because samples normally labeled *red* or *orange* in English are included in the category yellow when Vietnamese naming patterns establish the category boundaries. Note that the confidence rating peaks for monolingual speakers are largely unchanged; however, the bilingual Vietnamese speakers show two peaks, one at the optimal saturation level for yellow (as in Figure 1), and one at an optimal saturation and brightness level for orange. Figure 3 selected only those color stimuli named *orange* by English speakers. Note that despite the use of different category names (orange vs. dark yellow), confidence for monolingual English and Vietnamese speakers peaks at Level 2, but confidence for bilingual Vietnamese speakers peaks at Level 1. Because the bilingual Vietnamese and English groups selected the same high-confidence orange exemplar (OSA tile #246), we think this difference in the orange confidence peaks may result because the group includes individuals with divergent naming patterns assigning highest confidence to different exemplars. Monolingual Vietnamese speakers name *vang* (yellow) stimuli with higher confidence at three levels: (a) the optimal saturation level for yellow, (b) the optimal saturation level for orange, and (c) the optimal saturation level for red. Bigger differences in confidence are observed at optimal saturation levels for monolingual English and bilingual Vietnamese speakers, who draw on a less ambiguous category structure. These patterns are suggestive; however, it remains to be seen whether they can be replicated in other comparisons of bilingual and monolingual speakers.

The point of these analyses is that confidence seems to vary with several factors: (a) optimal saturation of the color stimuli judged, (b) existence of lexicalized category boundaries, and (c) existence of a consensual name that differentiates salient color appearances independent of category boundaries (e.g., within categories). The consensual name for a confidently named color need not be monolexemic and need not be a basic color term (using Berlin and Kay's 1969 definition of basicness). Confidence is higher when distinct names exist than when a single, modified category name must encompass varying color appearances. When a larger category includes several subcategories, existence of distinct names will produce peaks in confidence at the levels where the optimally saturated (most salient) exemplars matching those names are found. This can be seen in Figure 4 for the color *blue* (in English), which includes obvious subcategories for light and dark blue at

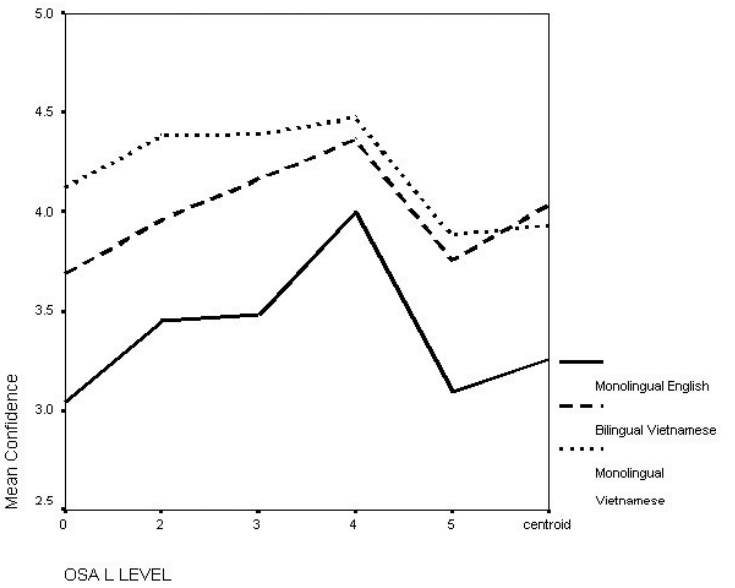

Figure 1: Mean Confidence Ratings for All Samples Classified as "Yellow" in English

different OSA L levels, in all three language groups (c.f. Paramei, 2005). This subdivision of the blue category exists in Vietnamese despite inclusion of green and blue within a single category (*xanh*) because modified and compound names exist to differentiate the color appearances within them.

The consequence of these observed differences in the naming patterns in two different ethnolinguistic cultures is that confidence ratings will vary with the naming patterns inherent to a particular language. Confidence ratings are not invariantly linked to the properties of color appearances, nor strictly connected to specific focal color stimuli extrapolated from pan-human neurophysiological response profiles; however, they are clearly affected by the naming practices of each language and culture. This means that confidence ratings may be used to determine the nameability of a color appearance and to make cross-cultural comparisons; however, such ratings will also be expected to vary with the naming practices of each culture. Thus, the qualities of nameability, high confidence in naming, and perceptual color salience cannot be

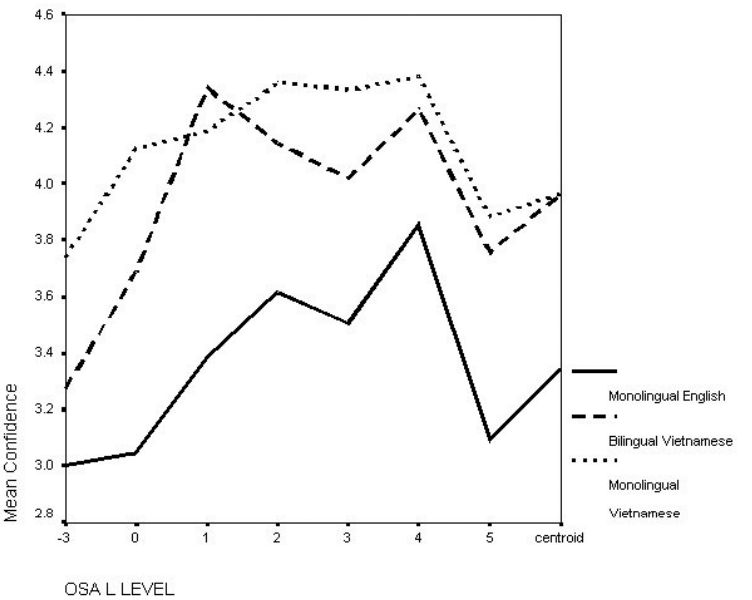

Figure 2: Mean Confidence Ratings for All Samples Classified *Vang* (Yellow) in Vietnamese

captured by the simple relationships inherent in the currently received view of color categorization and naming.

CULTURAL DIFFERENCES IN RATING CONFIDENCE

A danger of cross-cultural empirical investigation is that behavioral differences may arise from disparities in the cultural appropriateness of an empirical design for the assessed populations, or variation in the ways informants from different cultures respond or interact in the context of an empirical task. These complications must be considered when making direct comparisons of confidence ratings across cultures, especially where findings of overconfidence and underconfidence appear frequently and seem to vary with culture. Despite such considerations, one can find a number of investigations in the color categorization and naming literature (some discussed above) that liberally use culturally dependent cognitive measures as if they were culturally independent indices of

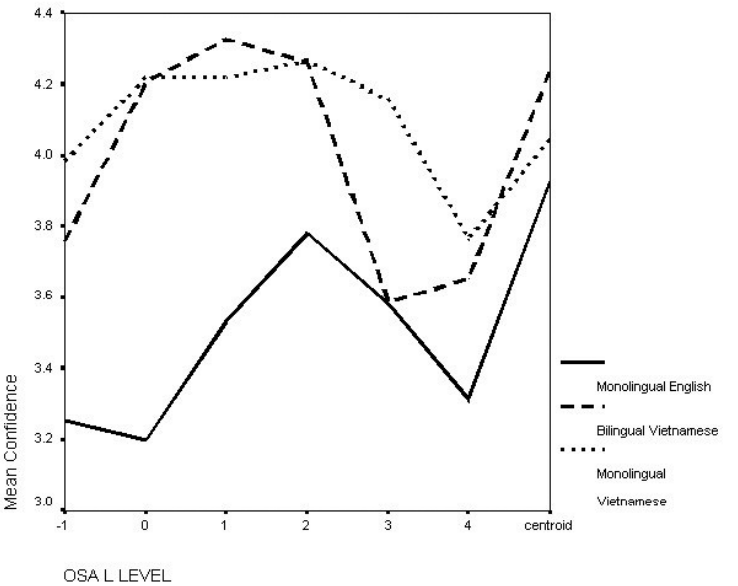

Figure 3: Mean Confidence Ratings for All Samples Classified as *Orange* in English

psychological processing that are appropriate for direct comparisons across cultures. The problem with such direct comparisons is evident in Figures 1 through 4. The mean confidence judgments of monolingual and bilingual Vietnamese speakers are higher than the mean confidence judgments of English speakers. This is consistent with what some have described as Asian "overconfidence" in general knowledge judgment contexts (Wright & Phillips, 1980; Wright et al., 1978). Where accuracy is concerned, overconfidence and underconfidence are miscalibrations of metaknowledge with respect to performance. Where objective accuracy can be assessed, Yates, Lee, and Bush (1997) defined *overconfidence* as a tendency to estimate a higher mean probability of correctness than the proportion of questions actually answered correctly. They reported that overconfidence is typically higher among Chinese than Western participant groups and asserted that this is not a data-analytic artifact (as has been suggested). The precise cause of this overconfidence is unclear. Yates, Lee, and Shinotsuka (1996) discussed several different explanations, critically evaluating the idea that self-esteem and other dispositional

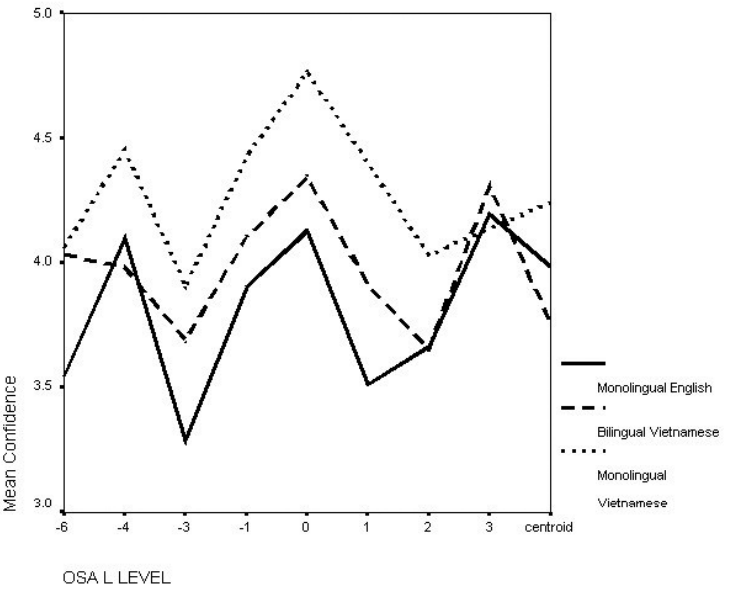

Figure 4: Mean Confidence Ratings for All Samples Classified *Blue* in English

traits mediate judgments of confidence across settings. As they noted, the basis for self-esteem may vary in collectivistic societies. They state: "In a collectivistic society, self-esteem is facilitated by adherence to the norm that one should 'fit in,' for example, that a person view him- or herself as having competence levels that are representative of the collective, not higher" (Yates, Lee, & Shinotsuka, 1996, p. 144). They argued that this implies that, with modesty, Asian overconfidence should be lower than, not higher than Western overconfidence. Lundeberg, Fox, Brown, and Elbedour (2000) found similar differences in confidence calibration comparing participants from Israel, the Netherlands, Palestine, Taiwan, and the United States, attributing high Palestinian overconfidence (compared to Israel) to a greater collectivism. They stated: "The Palestinian participants are more likely than Israeli or Western participants to adopt cognitive strategies designed to maintain a positive impression of themselves in the eyes of the community" (Lundeberg et al., 2000, p. 158).

Overconfidence also appears to vary depending on the type of judgment to be made (Lundeberg et al., 2000; Yates et al., 1997). It may be that collectivism is mediated less by esteem, whatever its basis, and more by understanding and attention to social and cultural norms relevant to a particular task. Individuals in collectivist cultures may be more strongly motivated, or more practiced, in referring to normative standards when making judgments. When a judgment task depends on a more normative comparison and depends less on absolute knowledge (such as whether an item was previously viewed or which line segment is longest), then cultural differences in confidence judgments may be enhanced. Cultural overconfidence has been found to change with the domain and judgment context (Lundeberg et al., 2000).

It may also be that overconfidence results from cultural differences in the use of rating scales, independent of what is being rated. Matsumoto (1994) noted a general disinclination among Asian participants to use the negative end of a scale, resulting in a rating bias. Baranski and Petrusic (1999) suggested that a cultural reluctance to use of the high end of the rating scale among Swedish participants (tested by Juslin & Olsson, 1997; Olsson & Winman, 1996) may account for observed underconfidence among Scandinavian participants. According to Baranski and Petrusic (1999), in many experiments, the Swedish participants never used the so-called certain category whereas their North American participants (in Ottawa and Toronto) used it 20% to 40% of the time. They concluded:

> Because people can be both over- and underconfident in both cognitive and sensory tasks, it is as problematic to speak of a general overconfidence bias in cognitive tasks as it is problematic to speak of a general underconfidence bias in sensory tasks. (Baranski & Petrusic, 1999, p. 1381)

If overconfidence occurs in some cultural groups of participants but underconfidence occurs in others, how good a measure can confidence be of a specific color appearance's cross-cultural nameability? To our knowledge, none of the previous investigators who have collected and analyzed confidence data for color-naming and categorization behaviors have attempted to address these problems with the cross-cultural intercomparability of such measures.

One general approach to dealing with systematic rating biases is to standardize (normalize) the ratings before comparing them. However, the quantitative transformation assumes that the rater is using the rating scale as a true interval scale (with equal distances between the values). It also assumes that different raters use the same anchor points, so that one person's 4 means the same as another person's 4. For confidence ratings, these are not valid assumptions, especially in an unexplored domain such as color naming. Standardization ignores these difficulties and makes it impossible to explore cross-cultural comparisons of differences in scale use. Nor are rating scales applied consistently across all trials—a fact obscured by the comparison of means. For example, it is well established that in general knowledge tasks, overconfidence is more likely with difficult questions and underconfidence is more likely with easy questions; however, does this same reversal occur with hard- or easy-to-discriminate sensory stimuli, much less hard or easy-to-name color appearances? Using a sensory discrimination task comparing horizontal line lengths, Baranski and Petrusic (1999) presented evidence that a transition from overconfidence to underconfidence occurs as accuracy decreases (at the 75% correct range, consistent with the transition point observed for cognitive judgments). It is possible, perhaps even likely, that confidence ratings are influenced by naming trial difficulty in a similar manner. These are issues that need to be considered when highly cognitive measures such as confidence are widely employed in investigations of cross-cultural color categorization and naming.

In our findings (Jameson & Alvarado, 2003), mean confidence was significantly different across the three language groups (monolingual English $M = 3.7$ on a scale from 1 to 5, bilingual Vietnamese $M = 4.1$, monolingual Vietnamese $M = 4.3$). If overconfidence exists in the Vietnamese ratings, there nevertheless exist stable relationships among the measures used in the current study, similar for both monolingual groups but different for bilinguals. Spearman's rho correlations between confidence and frequency and variability for the three language groups are shown in Table 2. Note that, despite higher confidence ratings (suggesting Asian overconfidence), the ratings of monolingual Vietnamese were more strongly correlated with measures of nameability than were bilingual Vietnamese speakers. This again suggests that monolingual participants in both language groups were more

responsive to the culture's normative naming practices than were bilingual participants and is consistent with our belief that confidence was mediated by this adherence to normative naming practices. The relationship between confidence judgments and nameability was highly similar in both monolingual groups, especially if confidence is correlated with a measure that incorporates frequency and variability into a single value. We created an agreement index by dividing the frequency by the variability. As shown in Table 2, correlations between confidence and this agreement index were more similar for the monolingual groups than for the bilinguals.

Based on these findings, we suggest that confidence ratings may not be directly comparable across cultures, because of cultural differences in mean confidence, perhaps resulting from general differences in the use of rating scales, but that use of well-conceived indirect measures of codeability or nameability may be justified. In particular, a comparison of the second-order within-language relationships between confidence, variability, and other nameability measures may be valid because such relationships appear consistent across the two groups of monolingual speakers in our study.

THE RELATIONSHIP BETWEEN CONFIDENCE AND RESPONSE TIME

Confidence and response time are not independent measures and thus do not provide two sources of confirmation of nameability, as proposed by Guest and Van Laar (2002). Baranski and Petrusic (2001) demonstrated in two experiments that the effects of requiring participants to make confidence judgments persist in trials continuing after participants are no longer required to make such judgments, suggesting that "some confidence processing occurs in parallel with the primary decision process" (p. 195). Furthermore, inclusion of a confidence judgment in a task increases decisional response times. Thus, response times analyzed in a study that includes confidence judgments are influenced by the processes required to produce confidence judgments and cannot be correlated with confidence in order to demonstrate nameability, as Guest and Van Laar (2002) wished to do.

Baranski and Petrusic (1999) argued that because "the properties of decision times in cognitive tasks mirror those observed in

sensory tasks" (p. 1381) a common decision process most likely operates in both domains. If so, the complexities of confidence judgments and response times found in sensory discrimination tasks, memory and general information tasks, or decision-making domains may apply to naming studies. For example, Vickers and Pietsch (2001) considered the contribution of the type of task, the cognitive demands on the participant, in determining the relationship between confidence and accuracy, with reference to Juslin and Olsson's (1997) sampling model of sensory discrimination. Although their consideration of task was very specific, they provided a detailed analysis of the underlying relationships among accuracy, response time, and confidence. In their analysis, they considered response time as a function of confidence, not solely confidence as a function of accuracy or response time. They warned:

> Without a sensitivity analysis of all parameters, there is always a danger that the behavior of a model is attributable not so much to the phenomenon under study but to the effect of "incidental" processing assumptions. . . . We find that the behavior of the sensory sampling model is heavily influenced by the assumptions of deadline responding, IOU [interval of uncertainty of the stimulus], and sequential sampling with delayed testing. (Vickers & Pietsch, 2001, p. 802)

In other words, task demands can and do influence observed relationships among confidence, response time, and accuracy of response. That these relations are not yet clearly understood in simple discrimination tasks implies that it is premature to assert the utility of either confidence judgments or response times to validate nameability in more complex color-naming tasks. Even so, the practice of using such measures has been popular since Boynton and Olson provided their first psychophysically oriented studies of basic color categories and naming (Boynton & Olson, 1987, 1990; Sturges & Whitfield, 1997). In general, it seems especially important to remember that tasks involving limitations on the stimulus set, the accessibility of names, or the quickness of response by participants will produce different results than tasks imposing different demands.

GUEST AND VAN LAAR'S STUDIES

Ultimately, Guest and Van Laar (2002) hoped to demonstrate the benefits of nameability in visual search tasks by showing improved accuracy, confidence, and response times for a set of highly nameable colors identified in their own previous research (Guest & Van Laar, 2000). They combined confidence ratings across three stimulus sets: (a) nameable stimuli identified in previous research as evoking consensual naming; (b) stimuli matched for Delta E differences to the most nameable samples based on "metric considerations"; and (c) stimuli selected for maximal discriminability during parallel visual search. Because the same ratings were used across the three stimulus sets, combining confidence ratings may seem permissible. However, working with means combined in this manner may be misleading, given the possibility that different task demands (including varying difficulty) and different stimulus sets may affect confidence ratings in different ways. For example, if more difficult choices result in overconfidence, can we conclude that the more nameable stimuli are actually easier to match, simply because they produce higher confidence ratings? Might the differences across the three stimulus sets actually be due to complex interactions between task difficulty, response time, and confidence that are different for each stimulus set?

Guest and Van Laar (2002) pointed out that perceptual differences (CIE Delta-E values; Wyszecki & Stiles, 1982) did not predict errors during their naming task. They suggested, on that basis, that use of nameable stimuli, associated with higher confidence ratings and named with faster response times, may produce fewer errors during a visual search task. Without actually testing nameable stimuli in a visual search task, this conclusion seems problematic. One cannot assume that high confidence necessarily correlates with accuracy. For example, Prinzmetal, Ivry, Beck, and Shimizu (2002) found that as confidence increases from 1 to 5 (on a scale from 1 to 9), the tendency to report illusory conjunctions also increases. If confidence increases still further from 5 to 9, fewer illusory conjunctions are reported and greater accuracy occurs. Thus, there is a region in which confident mistakes occur when performing a speeded-response task. However, this occurs only when stimuli have certain characteristics. Prinzmetal et al. (2002) noted that use of optimally saturated colors prevents such illusory conjunctions—their effect depends on the use of pastel colors.

Thus, false confidence itself may depend on stimulus characteristics. Furthermore, Delta E values may not predict errors during a task that is essentially normative because accuracy does not depend on stimulus differences but upon the fit between stimuli and names. In a visual search task, where accuracy is not based on a normative judgment, greater confidence may occur for more saturated colors and may have a stronger influence on the calibration of confidence judgments, without actually predicting accuracy. Thus, the relationship between optimal saturation, cultural naming practices, and confidence deserves further study. Without empirical confirmation, we think it is premature to assert that higher confidence and faster reaction times are properties of nameability resting upon cross-cultural universals.

IMPLICATIONS FOR THEORIES OF COLOR NAMING

Jameson (in press, 2005) proposed an alternative to the popular view that a pan-human color vision neural substrate produces universalities of color naming, resulting in the increased nameability of certain color appearances. She described a cognitivist view that hypothesizes separate mental representations for color percepts and semantic information about color. According to Jameson (2005), individuals maintain a perceptual representation of color based on their own perceptual capacities; however, they also maintain a shared cultural semantic representation of category structures and lexicalization specific to their culture. These two representations dictate separate and sometimes different color spaces and similarity structures that are linked by an additional level of cognitive representation, a cognitive naming function, that specifies the relations among the items in the two distinct spaces. It is this naming function that assigns names to color appearances and maps items in one space to items in the other space. These separate representations are most visible among people whose perceptual capacities are substantially different, such as dichromats (individuals who are color blind).

Jameson (in press, 2005) proposed that the universalities observed across cultures arise because pan-human cognitive processes are applied to solve common problems, such as the need to classify or differentiate items and communicate about them within the specific needs of a culture. Thus, the hierarchy of the emergence of basic terms identified by Berlin and Kay (1969), and the

deviations from it, can be explained by the application of the naming function under different conditions, the emphasis of different dimensions of color perception under different cultural pressures, and the availability and access to terms to name color appearances in different contexts and for different purposes.

Such a theory dictates that salience of certain regions of color space or certain highly saturated color appearances will be only one determinant of nameability. Others include availability of names within a culture's lexicon, extent of the stimulus set (the range and kind of color appearances to be named), and task demands, especially the need to make a speeded response. We have previously shown that relaxing the constraint for monolexemic naming and speeded response results in emergence of different most nameable stimuli in two different language groups (Jameson & Alvarado, 2003). That finding is problematic for Guest and Van Laar's (2000) method of identifying highly nameable exemplars using a codeability index sensitive to measures such as modal frequency and variability of naming. Here, we report that different exemplars can produce high confidence in different language groups, a finding that further supports Jameson's (in press, 2005) proposed model of naming and does not support the view that high confidence arises from a special salience determined by panhuman neural visual processing.

More specifically, our findings suggest that in a hue-based naming system such as English, where naming patterns include a preference for use of monolexemic terms (contextualized as well as basic), the saturation of that hue contributes to increased confidence in naming. In a language such as Vietnamese, where polylexemic terms are preferred and modifiers are used to describe lightness (brightness) in addition to hue, consensually named best exemplars varying in lightness are named with highest confidence, as well as a few of the same highly saturated exemplars highly nameable in English. We speculate that a language that emphasized brightness before hue (as Vietnamese does not) would produce a list of most nameable stimuli considerably different, perhaps ignoring saturation.

We believe our findings support the idea that confidence is a sensitive measure of the goodness of fit between a lexical term and its referent, a specific color appearance. If we consider nameability to be the propensity for consensual naming, confidence may indicate the extent to which a person feels he or she has adhered to cultural naming practices when naming a sample. Thus, confidence

should vary with people's sensitivity to cultural norms and their inclusion within that culture. The comparison of bilingual speakers with monolinguals suggests that confidence is a poorer index for those who are less "tuned-in" to an ethnolinguistic culture's naming patterns. If nameability were solely linked to pan-human color vision, then confidence should not vary with cultural inclusion, as it did in the current study. Nor should confidence vary with the category boundaries and naming of yellow and orange, as it did in the current results.

Our enumeration of the complexities of using confidence to assess nameability is meant to suggest that confidence rating averages cannot be directly compared across cultures without taking into account a variety of factors. We believe that confidence ratings of the goodness of a name evaluate the quality of mapping performed using the naming function. Confidence ratings made in the context of other tasks may evaluate different things and thus may not be directly comparable. In fact, they may be only tenuously related to any property of nameability. Guest and Van Laar (2002) may be entirely correct that use of highly nameable stimuli may improve accuracy in visual search tasks. We think that remains to be demonstrated and are skeptical that highly nameable stimuli will be the same across cultures with widely different naming patterns. This implies that the kind of stability and wide usability Guest and Van Laar (2002) sought may not be attainable, although using nameable stimuli may provide the closest approach to it.

Similarly, our position regarding the hotly debated views of color categorization universals and color-naming relativism is that of measured acceptance of both. In general, we consider that color categorization and naming within and across cultures are as much influenced by constraints and commonalities imposed by visual-processing mechanisms as they are influenced by common culturally learned naming practices and the similar ways cultures come to form their naming systems.

References

Alvarado, N., & Jameson, K. A. (2002). The use of modifying terms in the naming and categorization of color appearances in Vietnamese and English. *Journal of Cognition and Culture*, *2*(1), 53-80.

Baranski, J., & Petrusic, W. (1999). Realism of confidence in sensory discrimination. *Perception and Psychophysics*, *61*, 1369-1383.

Baranski, J., & Petrusic, W. (2001). Testing architectures of the decision-confidence relation. *Canadian Journal of Experimental Psychology*, *55*, 195-206.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Boynton, R., & Olson, C. (1987). Locating basic colors in the OSA space. *COLOR Research and Application*, *12*, 94-105.

Boynton, R., & Olson, C. (1990). Salience of chromatic basic color terms confirmed by three measures. *Vision Research*, *30*, 1311-1317.

Guest, S., & Van Laar, D. (2000). The structure of colour naming space. *Vision Research*, *40*, 723-734.

Guest, S., & Van Laar, D. (2002). The effect of name category and discriminability on the search characteristics of colour sets. *Perception*, *31*, 445-461.

Hardin, C. L. (2005). Explaining basic color categories. *Cross-Cultural Research*, *39*, 72-87.

Hardin, C. L., & Maffi, L. (1997). *Color categories in thought and language*. Cambridge, UK: Cambridge University Press.

Ishihara, S. (1987). *Ishihara's tests for colour-blindness* (Concise Edition). Tokyo: Kanehara.

Jameson, K. A. (in press). Culture and cognition: What is universal about the representation of color experience? *Journal of Cognition and Culture*.

Jameson, K. A. (2005). Why GRUE? An interpoint-distance model analysis of composite categories. *Cross-Cultural Research*, *39,* 159-204.

Jameson, K. A., & Alvarado, N. (2003). Differences in color naming and color salience in Vietnamese and English. *COLOR Research and Application*, *28*(2), 113-138.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review*, *104*, 344-366.

Kay, P., & Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist*, *101*, 743-760.

Kuehni, R. (2004). Variability in unique hue selection: A surprising phenomenon. *COLOR Research and Application*, *29*, 158-162.

Lundeberg, M., Fox, P., Brown, A., & Elbedour, S. (2000). Cultural influences on confidence: Country and gender. *Journal of Educational Psychology*, *92*, 152-159.

MacLaury, R. E. (1992). From brightness to hue: An explanatory model of color category evolution. *Current Anthropology*, *33*, 137-186.

Matsumoto, D. (1994). *Cultural influences on research methods and statistics.* Belmont, CA: Brooks/Cole.

Olsson, H., & Winman, A. (1996). Underconfidence in sensory discrimination: The interaction between experimental setting and response strategies. *Perception and Psychophysics*, *58*, 374-382.

Paramei, G. (2005). Singing the Russian blues: An argument for culturally basic color terms. *Cross-Cultural Research 39*, 10-38.

Prinzmetal, W., Ivry, R., Beck, D., & Shimizu, N. (2002). A measurement theory of illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 251-269.

Sturges, J., & Whitfield, T. W. A. (1997). Salient features of Munsell colour space as a function of monolexemic naming and response latencies. *Vision Research*, *37*(3), 307-313.
Vickers, D., & Pietsch, A. (2001). Decision making and memory: A critique of Juslin and Olsson's (1997) sampling model of sensory discrimination. *Psychological Review*, *108*, 789-804.
Webster, M. A., Webster, S. M., Bharadwaj, S., Verma, R., Jaikurnar, J. Madan, G., et al. (2002). Variations in normal color vision. III. Unique hues in Indian and United States observers. *Journal of the Optical Society of America A*, *19*, 1951-1962.
Wright, G., & Phillips, L. (1980). Cultural variations in probabilistic thinking: Alternative ways of dealing with uncertainty. *International Journal of Psychology*, *15*, 239-257.
Wright, G., Phillips, L., Whalley, P., Choo, G., Ng, K., Tan, I., et al. (1978). Cultural differences in probabilistic thinking. *Journal of Cross-Cultural Psychology*, *9*, 285-299.
Wyszecki, G., & Stiles, W. (1982). *Color science: Concepts and methods, quantitative data and formulas.* New York: John Wiley
Yates, J. F., Lee, J.-W., & Bush, J. (1997). General knowledge overconfidence: Cross-national variations, response style, and "reality." *Organizational Behavior and Human Decision Processes*, *70*, 87-94.
Yates, J. F., Lee, J.-W., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes*, *65*, 138-147.

Nancy Alvarado has a Ph.D. in psychology from the University of California, Irvine. She is an assistant professor of psychology at California State Polytechnic University, Pomona. Her research interests include the categorization and naming of emotional facial expressions, emotional states, and color appearances. She is most interested in similarities between emotion naming and color naming across cultures.

Kimberly A. Jameson has a Ph.D. in psychology from the University of California, Irvine. Her research seeks to understand the ways individuals and groups conceptualize naturally occurring categories in the real world. This includes empirical studies of color categorization and naming; the modeling of concept formation for perceptual stimuli (e.g., individual cognitive organization of color sensations and its relation to linguistic classifiers); evaluation of performance in tasks involving color-coded information; and identifying cross-cultural cognitive universals of color appearance.