

# Even with a green card, you can be put out to pasture and still have to work: Non-native intuitions of the transparency of common English idioms

BARBARA C. MALT

*Lehigh University, Bethlehem, Pennsylvania*

and

BRIANNA EITER

*State University of New York, Binghamton, New York*

Native speakers of English use idioms such as *put your foot down* and *spill the beans* to label events that are not described literally by the words that compose the idioms. For many such expressions, the idiomatic meanings are transparent; that is, the connection between the literal expression and its figurative meaning makes sense to native speakers. We tested Keysar and Bly's (1995) hypothesis that this sense of transparency for the meaning of everyday idioms does not necessarily obtain because the idiomatic meanings are derived from motivating literal meanings or conceptual metaphors, but rather (at least in part) because language users construct explanations after the fact for whatever meaning is conventionally assigned to the expression. Non-native speakers of English were exposed to common English idioms and taught either the conventional idiomatic meaning or an alternative meaning. In agreement with Keysar and Bly's suggestion, their subsequent sense of transparency was greater for the meaning that the speakers had learned and used, regardless of which one it was.

Idioms label events that are not described literally by the words that compose them. Thus speakers of English use the phrases *skating on thin ice*, *spilling the beans*, and *adding fuel to the fire* to mean doing something dangerous, accidentally telling a secret, and making a situation worse, respectively. The figurative or nonliteral meanings for many idioms are considered transparent; that is, the connection between the literal expression and its figurative meaning makes sense to native speakers (Keysar & Bly, 1995).

This transparency has important implications for theories of how idioms are represented in the mental lexicon and how they are used in production and comprehension. Linguists traditionally have taken idioms to be noncompositional—that is, to have meanings that do not derive from their constituent words—and so have treated them as unitary lexical items. Following this tradition, some psycholinguistic theories have considered idioms to have separate entries in the mental lexicon (e.g., Bobrow &

Bell, 1973; Swinney & Cutler, 1979). But this sort of approach does not take into account the intuitions of transparency that speakers have for many common idioms.

More recent approaches to idioms are compatible with intuitions of transparency and indeed motivated in large part by them. One such approach takes the meanings of idioms to be grounded to some extent in the meanings of their constituent words (e.g., Cacciari & Glucksberg, 1991; Glucksberg, 1991; Nunberg, Sag, & Wasow, 1994). For instance, the meanings of *kick* and *bucket* constrain use of the expression *kick the bucket*; one can lie dying all week but one cannot lie kicking the bucket all week, because dying is a continuous act and kicking a bucket is a discrete act (Glucksberg, 1991). Another takes the meanings of idioms to derive from the relation of elements of the idiom to independently existing conceptual metaphors. So, for instance, the idiom *keep someone at arm's length* is understood in terms of the metaphors INTIMACY IS PHYSICAL CLOSENESS and SOCIAL (OR PSYCHOLOGICAL) HARM IS PHYSICAL HARM (Lakoff, 1987, p. 448; see Keysar & Bly, 1999). In either case, the sense of transparency is taken to derive from the connection between the motivating words or metaphors that are hypothesized to contribute to the creation of that meaning and the understood meaning of the idiom.

Keysar and Bly (1995) proposed, however, that at least part of the sense of transparency for the meanings of common idioms does not arise because the idiomatic mean-

---

This research was carried out as an honors project submitted by B.E. to the Department of Psychology at Lehigh University. It was supported by a Lehigh Forum Student Research Grant to B.E. and NIMH Grant MH51271 to B.C.M. and Steven Sloman. We thank Cindy Connine and Alexandria Guzman for helpful comments on a previous draft of this article. Please address correspondence to B. C. Malt, Department of Psychology, 17 Memorial Drive East, Lehigh University, Bethlehem, PA 18015 (e-mail: barbara.malt@lehigh.edu).

ings are understood through reference to motivating literal meanings or conceptual metaphors. Rather, after people learn the conventional meaning of an idiom, they construct links between the idiomatic meaning and the words of the phrase, generating an explanation for themselves of why the phrase means what it does. Once these links have been created and strengthened through repeated use, it becomes difficult to construct other meanings or explanations for how those alternative meanings would be linked to the expression.

In illustration, consider the idiom *keep someone at arm's length*. Neither the words that constitute the idiom nor the possible existence of a conceptual metaphor INTIMACY IS PHYSICAL CLOSENESS dictates that the expression *must* mean "to keep someone psychologically or socially distant." It could, in fact, reasonably refer to keeping someone nearby (no farther away than an arm's length; never out of reach). Once the conventional interpretation of the expression is learned, however, and an explanation is constructed in terms of a metaphor of physical distance, the relation between the two seems obvious and the possibility of a different meaning seems implausible.

To test their hypothesis, Keysar and Bly (1995) presented obsolete idioms such as *the goose hangs high*, *lay out in lavender*, and *row cross-handed* to native speakers of English. The participants were not familiar with the idioms and so did not have a preestablished sense of transparency for any particular meanings in connection with the idioms. In their first experiment, Keysar and Bly used three possible meanings for each idiom: the original, a reversed meaning, and an unrelated meaning. For instance, for the idiom *to have someone dead to rights*, the original meaning is "to catch someone in the act, unquestionably guilty." The reversed meaning that they constructed was "to convict someone in spite of unquestionable innocence," and the unrelated meaning was "to give someone an award." Participants read paragraphs containing the idioms that were designed to bias them to one of the three meanings and indicated which they thought the meaning of the idiom was. They were then asked to consider a person who encounters the idiom in an uninformative context (e.g., by reading a newspaper headline, "The City Council Has the Mayor Dead to Rights") and to judge which of the three meanings the person would think the idiom had. Keysar and Bly found that participants tended to attribute to the person whichever interpretation they themselves had learned.

In a second experiment, participants first gave baseline transparency judgments of two possible meanings for each idiom and then read the idioms in paragraphs designed to bias them to either the original or a reversed meaning. They were asked to identify each idiom's meaning, to copy the meaning, and, for a third of the items, to use the idiom in a sentence. For another third of the items, they used the idiom in two sentences; for the remaining third, they did not construct any sentences. The participants were then given both meanings for each idiom again and were asked to rate the extent to which each meaning made sense. Keysar and Bly (1995) found that the ratings for learned

meanings in the test phase were higher than the ratings for the same meanings in the baseline task, and the meanings that had been learned (whether original or reversed) were rated as more sensible in the test phase than the meanings that had not been learned. Furthermore, the more times a meaning had been used (in the sentence construction task), the less sensible the other possible meaning was rated. These results support their hypothesis that a sense of transparency for the meanings of idioms is at least in part a consequence of learning a particular meaning for an idiom, rather than being driven a priori by a motivating connection between the words of an idiom or conceptual structures and the conventional meaning.

Although Keysar and Bly (1995) used obsolete idioms in order to be able to manipulate the learned meaning for native speakers of English, it is crucial to their hypothesis that the results generalize to common idioms in current use. Keysar and Bly (1999) provide arguments against the possibility that their earlier results came about only because the conceptual structures associated with the obsolete idioms have also died, leaving participants more susceptible to suggestion about the nature of the meanings. However, it remains a possibility that the obsolete idioms themselves in some way differ in nature from idioms currently in common use. For instance, given that the vocabulary words involved in the obsolete idioms remain familiar today and the concepts that the meanings capture do also, yet the idioms have disappeared from the language, one might speculate that the original motivation for these idioms was particularly weak or obscure, leaving them with reduced staying power and less inherent transparency. In addition, the fact that their native speaker participants would have recognized that these expressions are not part of the English language as they know it, and might guess that the experimenters invented them for purposes of the study, could have induced experiment-specific response strategies such as making a deliberate effort to construct an explanation for the meaning given in order to rationalize the experimenters' choice.

The primary purpose of the present study was to test Keysar and Bly's hypothesis about how a sense of transparency for idiomatic meanings develops using common idioms that are in current use in American English. To do so, we presented the idioms to non-native speakers of English whose grasp of English was sufficiently good that they were familiar with the vocabulary in the idiomatic expressions (and, in fact, sufficient to function in a university setting and to follow fairly complex experimental instructions), but who had little or no familiarity with the idioms used. Following Keysar and Bly's (1995) second experiment, baseline judgments of transparency were first collected for each of two meanings: the true idiomatic meaning or an alternative meaning. The idioms were then presented in biasing paragraphs that would lead participants to understand one of the two meanings. Participants were asked to select the meaning from two choices and then to write several sentences using the phrase appropriately with the selected mean-

ing. In a subsequent test phase, the participants received all the idioms presented during the learning phase and were asked again to rate the extent to which each meaning made sense. Our expectation is parallel to that of Keysar and Bly (1995): Whatever idiomatic meaning the non-native speakers of English are exposed to will, through learning and use, become the transparent idiomatic meaning.

A secondary purpose of this experiment was to provide converging evidence for the learning effect that would help rule out any possibility that the effect arises due to demand characteristics of the experimental situation. Both Keysar and Bly's transparency rating instructions and ours emphasized to participants that there may be more than one meaning that makes sense for an idiom, and participants were expected to judge how much sense each meaning makes independent of the other sense. Keysar and Bly report evidence against the possibility that participants are inclined to make their ratings contrastive—giving a low rating to one meaning simply because they have given a higher rating for the other—by showing that the ratings for the two meanings of each idiom were not negatively correlated in the baseline rating task (and were, in fact, modestly positively correlated). However, this observation does not address the possibility that in the test phase, participants responded to what they took the experimenter's expectation to be and said that whatever meaning they had just been taught was the one that made the most sense. In order to more directly probe participants' ability to associate the correct meaning with the idioms, after they completed the transparency ratings, we told the participants that they might have learned a false meaning for some of the idioms, and we asked them to go back through the list of idioms and choose the meaning that they thought was the true idiomatic meaning. If a tendency to prefer the learned meaning is due to an actual perception of greater transparency, rather than experimental demand, the participants should demonstrate the learning effect in this measure as well.

## PRETEST

A pretest was used to select the idioms that would be used in the next part of the experiment. The goal was to choose a set of idioms that was unfamiliar to non-native speakers of English from among those in current use in American English. The pretest also provided an opportunity to identify appropriate language qualifications for participants for the rest of the experiment.

## Method

**Participants.** Nine non-native speakers of English participated, consisting of 7 Lehigh University undergraduates, 1 graduate student, and 1 professor.

**Materials.** Fifty-three idioms from the *Longman Dictionary of English Idioms* (1979) were selected that referred to actions and that are in current use in American English discourse (based on the authors' judgment). For example, some of the idioms were *rub salt into his wounds*, *avoid like the plague*, and *jump the gun*. The 53 idioms were presented to the participants on a questionnaire. For each

idiom, participants were asked to answer "yes" or "no" to two questions: "Have you heard this before?" and "Do you have an idea of the meaning?" Following the idioms section, the questionnaires asked about the participant's language history, including what their native language was, at what age they had begun to learn English, and how many years they had been living in an English-speaking environment.

**Procedure.** The participants were given the questionnaires and were simply asked to answer all questions to the best of their ability. They were told that there were no right or wrong answers.

## Results

The number of "no" responses was tallied for each of the two questions. The idioms that participants most frequently indicated they had not heard before and did not know the meaning of were chosen for the stimuli set for the next phase of the experiment. For the "heard before" question, the idioms chosen received "no" responses from at least 7 participants, and for the "have an idea of the meaning" question, the idioms chosen received "no" responses from at least 5 participants. Eighteen idioms met these two criteria, with a mean of 7.89 "no"s for "heard before" and 6.44 "no"s for "have an idea of the meaning". These idioms are given in the Appendix.

The information collected from the language history questionnaire was used to identify characteristics of participants for the main experiment who would have low familiarity with the idioms. In particular, it appeared that participants who had lived in an English speaking environment more than 10 years would be likely to know many of the idioms and so would not be appropriate participants.

## EXPERIMENT

### Session 1

The main experiment consisted of two sessions. Session 1 was conducted both to further narrow down the idioms to be learned in Session 2 and to select the participants who would participate in Session 2. It also provided baseline transparency ratings to compare with Session 2 ratings. The participants first provided information about their language history. They then took a vocabulary test to determine their knowledge of the words found within the idioms, so that we could eliminate any participants whose English vocabulary was not sufficient to have the potential to construct explanatory links between the idioms and the to-be-presented meanings. Next the participants rated their familiarity with the 18 idioms chosen from the pretest in order to allow us to eliminate any participants who were already familiar with most of the idioms. Finally, following Keysar and Bly (1995), the participants gave judgments of how much sense they felt each of two meanings made for each idiom in order to provide baseline transparency ratings.

### Method

**Participants.** The participants were 32 non-native speakers of English. Twenty-one were drawn from the introductory psychology participant pool and participated for course credit. Eleven were recruited

from the Lehigh community and participated for a monetary reward. All were undergraduate or graduate students at Lehigh University.

**Materials.** The language history questionnaire asked the participants for their age, birth place, native language, other languages spoken, the age at which they began to speak English, how many years of formal education in English they had had, and how many years they had been living in an English-speaking environment.

The vocabulary test contained 19 vocabulary words consisting of the content words in the 18 idioms that had a frequency of less than 200 according to Kučera & Francis (1967). Each vocabulary word was followed by four answer choices: one correct choice describing the meaning of the word or a location or a characteristic closely related to its meaning, and three incorrect choices that had other meanings of differing degrees of semantic relatedness. For instance, for the word *grin*, the choices were “cry,” “smile,” “frown,” and “teeth”; for the word *mouth*, they were “part of the body found on the leg,” “part of the body found on the face,” “part of the body found on the neck,” and “part of the body found on the back.”

The familiarity rating questionnaire contained the 18 idioms chosen from the pretest (given in the Appendix), preceded by instructions (see below) and an 8-point familiarity scale. The scale was labeled with 0 as *never seen before*, 1 as *slightly familiar with the idiom and its meaning*, 4 as *moderately familiar with the idiom and its meaning*, and 7 as *very familiar with the idiom and its meaning*.

The transparency rating questionnaire listed each of the 18 idioms along with the true idiomatic meaning for each, taken from the *Longman Dictionary of English Idioms* (1979), and an alternative that we constructed. The alternatives were mostly opposites (e.g., for the idiom *break the bank*, the possible meanings were “win all the money that is being risked at a game of chance,” and “lose all the money that is being risked at a game of chance”); in a few cases they were not opposites but were nevertheless distinctly different (e.g., for the idiom *to make their hair stand on end*, the true meaning was “to cause someone to feel great fear” and the alternative was “to make someone look scary”). The two meanings were listed to the right of each idiom. Two different forms of the rating sheet were used, each with the 18 idioms in the same random order but counterbalancing the order of presentation of the two meanings for each. No indication was given of which meaning was the true meaning (nor whether either of them was, in fact, the true meaning). Instructions were given at the top of the first page (see below), followed by an 8-point transparency scale. The scale was marked with 0 as meaning *makes no sense*, 1 as *small amount of sense*, 4 as *medium amount of sense*, and 7 as *large amount of sense*.

**Procedure.** Participants were run individually in 20-min sessions. They first completed the language history questionnaire, then the vocabulary test, then the familiarity ratings, and finally the transparency ratings. The participants were told that this experiment was testing non-native speakers’ knowledge of idioms, that there were several parts to the experiment, and that for each of the parts, there were no right or wrong answers. The participants were simply asked to respond to the language history questions. Each subsequent set of directions was presented in writing to the participants and read aloud by the experimenter while the participant followed silently.

For the vocabulary test, the participants were instructed to “circle the word or phrase that best defines or describes the meaning of, location, or characteristic that is most closely related to the bolded words found below.” For the familiarity ratings, they were simply told, “Using the scale from 0 to 7 below, please rate how familiar you are with the following idioms.”

For the transparency ratings, the participants were told,

Consider the phrase *to have your back against the wall*. The true definition of this idiom is “to be in a very difficult situation in which you have to defend yourself,” but there are other possible definitions that could make sense in relation to the phrase. For instance, this idiom

could mean “to be physically unable to move away from a wall because you are in a certain space,” or “to be unable to hide or protect yourself from harm because there is no shelter.” Either of these definitions could be the true meaning of the idiom *to have your back against the wall* because they seem to make sense, or relate, to the words that make up the idiom.

The participants were then instructed to use the scale that followed to rate how much sense each of the two meanings given made for each idiom.

Once the participants were done with their ratings, they were told that they would be notified in several days whether or not they were to continue to participate in the experiment, and that until that time, they should not discuss the idioms with anybody else. Participants who were not asked to continue were mailed a debriefing form and given their course credit or payment. Participants who were asked to continue were contacted by phone or e-mail to schedule the second session.

## Results

The vocabulary test was administered to determine whether participants had adequate knowledge of the vocabulary words in the idioms. Participants who correctly identified the meanings for at least 80 percent of the vocabulary words qualified for further consideration for Session 2. All 32 participants who took part in Session 1 met this criterion.

The familiarity ratings were also examined to determine which participants would be asked to participate in Session 2. Participants who gave ratings below 4 on the 8-point rating scale to more than half of the selected idioms were asked to participate. Twenty-four of the 32 participants met this criterion. Since all had passed the vocabulary test, these 24 were asked to continue. The language demographics of the participants who continued, derived from the language history questionnaire, will be reported in the context of Session 2.

Mean transparency ratings were calculated for the two meanings of each idiom and were used to select idioms for Session 2. Ideally, idioms with about equal mean transparency ratings would be used in case the level of transparency affects susceptibility to shifting with learning. It is also preferable for the ratings to fall at about the midpoint of the scale so that both increases and decreases in transparency as a function of learning and use can be readily detected. However, we apparently were not successful at generating alternative meanings that were a priori seen as being as transparent as the true meanings to this group: The mean transparency rating for the true meanings was 5.11 and for the alternatives was 2.47. To avoid potential floor effects for the alternative meanings, we selected the eight idioms that had the highest mean transparency ratings for the alternatives. The idioms selected are indicated with an asterisk in the Appendix. The alternative meanings for these eight all had ratings of 2.65 or higher, with a mean of 3.17. The true meanings for these idioms had ratings of 3.61 or higher, with a mean of 5.00. Thus, there was a modest imbalance a priori favoring the transparency of the true meanings. Note, however, that for the subset of participants who participated in Session 2, the ratings were closer to equal

(as reported below; see Table 2), suggesting that the imbalance was due mainly to a higher level of familiarity with the idioms by the larger group.

## Session 2

The second session of the main experiment used the eight idioms chosen in Session 1 to test the hypothesis that whatever implied idiomatic meaning non-native speakers of English are exposed to will, with subsequent learning and use, become the transparent idiomatic meaning. Participants read biasing paragraphs, selected the meaning for the idiom that they felt the paragraph implied, and then copied the selected meaning and wrote two sentences of their own using it. They then rated both meanings for all the idioms for transparency and, finally, made judgments of which was the true meaning.

## Method

**Participants.** Nineteen of the 24 non-native speakers of English who qualified from Session 1 participated in this experiment. The remaining 5 declined to participate. Eleven were drawn from the introductory psychology participant pool and participated for course credit. The remaining 8 were from the larger university community and participated for monetary reward. The participants came most heavily from Asian language backgrounds (4 Chinese, 3 Korean, and 2 Japanese). Other languages represented were Spanish (4), Turkish (2), French (1), Tamil (1), Greek (1), and Indonesian (1). The participants had begun to learn English at a mean age of 11.84 (median = 12) and had been in an English-speaking environment for a mean of 5.03 years (median = 4).

**Design.** The independent variable was the meaning learned for each idiom during the learning and use phase: either the true idiomatic meaning or the alternative meaning that we had constructed. This variable was manipulated within subjects: Each person received true meanings for half the idioms and alternative meanings for the other half. The main dependent variable was the transparency rating. An additional dependent measure was the meaning that participants circled when asked to identify which meaning was the true one.

**Materials.** The eight idioms selected in Session 1 formed the basis for the materials. The mean familiarity rating for the eight selected idioms given by the Session 2 participants in Session 1 was 1.69 ( $SD = 1.1$ ) on the 8-point scale. Sixteen biasing paragraphs were constructed, 2 for each of the eight idioms, with 1 paragraph biasing for the true idiomatic meaning and the other for the alternative meaning. The paragraphs were all approximately 100 words long and used words and concepts likely to be familiar to non-native speakers of English. The idiom within each was underlined.

The paragraphs were presented in booklets. Each page contained one paragraph, along with two answer choices for each of the eight idioms. Half of the paragraphs in each booklet were those biasing interpretation of the meaning toward the true idiomatic meaning, and half were those biased toward the alternative meaning. For each paragraph, the biased meaning was one answer choice; the other answer choice was an unrelated meaning (in order to avoid exposing participants again to the other meaning before giving their second transparency ratings). To create foils that had sufficient links to the paragraphs to require the participants to think carefully about which meaning was correct, the unrelated meanings were not always the same for the two paragraphs for each idiom. For instance, for the idiom *to run rings around*, in one version, the biasing paragraph was about a football rivalry and the meaning choices were “to be a great deal better at” (the true meaning) and “to draw a circle around a team” (the unrelated meaning); in the other version, the biasing paragraph was about two sisters with different degrees of athletic ability and the choices were “to be much worse at something than

someone else (the alternative meaning) and “to be able to run very fast” (the unrelated meaning). There were two versions of the booklet, with meaning type (true vs. alternative) counterbalanced so that each idiom appeared in half the booklets with its true meaning and in half with its alternative meaning, and each participant was exposed to four true meanings and four alternative meanings. Table 1 gives an example of an idiom with each of its two biasing paragraphs and response options. The order of the idioms was randomized in each booklet by shuffling the pages before stapling. Each paragraph was followed by three lines labeled “1,” “2,” and “3” for use as described below.

The transparency rating test was the same as that administered during Session 1. To ensure that the context of the ratings remained the same as it was in Session 1, the participants gave transparency ratings for both meanings for all 18 idioms—the 8 learned and the remaining 10 from Session 1.

**Procedure.** Participants were run individually in sessions that lasted approximately 1 h. First participants learned a meaning for each idiom in a procedure based on Keysar and Bly’s (1995) Experiment 2. The participants were presented with the booklets containing the biasing paragraphs for each idiom and asked to choose the meaning of the idioms found within the paragraphs from the two choices. Participants were told that, for example, after reading a paragraph with the idiom *to grin and bear it* underlined within it, they should circle the answer choice that best described the meaning of the idiom within that paragraph. They completed a sample paragraph to ensure comprehension of the instructions. Once the participants finished reading all paragraphs and choosing the meaning of the idioms, they were asked, for each idiom, to recopy their selected idiomatic meaning on the line labeled “1” and then to create two of their own sentences using the idiom on the lines labeled “2” and “3.” This exercise was to ensure that participants learned well and would remember the idiomatic meaning. They were told that the sentences they created could be simple, as long as they conveyed the meaning of the idiom, and they were given the example, “The woman who kicked the bucket was buried last week.”

After completing this phase, the participants handed in their booklets and received the transparency rating sheet. The transparency rating instructions were the same as in Session 1.

After completing the transparency ratings, the participants were instructed orally that the meanings they learned during the session might not have been the true meaning for each idiom, and they were asked to go back through this rating sheet and circle the meaning that they felt was the true meaning for each idiom.

**Table 1**  
**Sample Biasing Paragraphs and**  
**Response Options for Session 2**

### Version A

Steve’s birthday is in three weeks. He will turn 65 years old on that day. Steve has been working for the same company for almost 45 years and does not wish to stop. Unfortunately, the new management wants younger employees and has asked Steve to retire. Steve is being put out to pasture because of his age.

- A. To give an award to someone
- B. To make a person stop working because of old age

### Version B

Because there was such a need for the new product and not enough people who were the right age to work in the factory, the manager went into the community to find workers. He had to ask younger children and elderly people to work in the factory and make the product. The manager put this group out to pasture regardless of their ages.

- A. To put somebody to work, regardless of age
- B. To ask someone for money

**Table 2**  
**Mean Transparency Ratings for Not-Learned and Learned Meanings in Session 1 and Session 2**

	Learning	
	Not Learned	Learned
Session 1		
True meaning	4.47	4.84
Alternative meaning	3.20	3.24
Session 2		
True meaning	2.75	6.79
Alternative meaning	1.39	5.95

## Results

We had hypothesized that whatever idiomatic meaning the non-native speakers of English were exposed to—whether the true idiomatic meaning or an alternative—would, with subsequent learning and use, become the transparent idiomatic meaning. The transparency ratings were analyzed to determine whether participants felt the meanings they learned were more transparent than the meanings they did not learn. Baseline data for comparison were the ratings for the 8 learned idioms, extracted from the complete Session 1 data, given by the 19 participants who continued on to Session 2. (Meanings referred to as “learned” and “not learned” for Session 1 are the meanings that would *become* the learned and not-learned meanings for each participant in Session 2; no meanings were actually learned in the first session.) The ratings for each session, for both meanings when learned and when not learned, are given in Table 2.

An overall 2 (session: 1 vs. 2)  $\times$  2 (learning: learned vs. not learned)  $\times$  2 (meaning: true vs. alternative)  $\times$  2 (booklet version: 1 vs. 2) analysis of variance (ANOVA) was carried out with the first three factors as within subjects and booklet version (reflecting the two counterbalanced stimulus sets) as between subjects. The ANOVA showed a significant effect of learning [ $F(1,17) = 96.37, p < .0001$ ], reflecting higher ratings for learned items than for not-learned items, and a significant interaction between session and learning [ $F(1,17) = 71.47, p < .0001$ ], reflecting little difference between learned and not-learned items in Session 1 and a large difference in Session 2. The main effect of session itself was not significant [ $F(1,17) = 2.88, p > .10$ ], due to the interaction with learning (with learned items being rated higher and not-learned ones lower in Session 2). The effect of booklet version was not significant [ $F(1,17) = 2.99, p > .10$ ], nor was its four-way interaction with session, learning, and meaning [ $F(1,17) = .45, p > .50$ ], indicating that the effects were similar for both stimulus sets. Booklet version did interact significantly with learning [ $F(1,17) = 5.3, p < .05$ ] and with learning and meaning [ $F(1,17) = 12.61, p < .005$ ], due to a somewhat reduced effect of learning across sessions for true meanings for the second booklet version. However, a separate ANOVA for this booklet alone showed that the main effect of learning remained significant for this version [ $F(1,9) = 33.22, p < .001$ ] as did the interaction of session with learning [ $F(1,9) = 36.05, p < .001$ ]; thus we collapsed over the two booklet versions

in further analyses. These and all results below were fully paralleled in analyses across items.

The effect of learning was analyzed in more detail by comparing the transparency ratings for the learned and not-learned meanings from Session 2 with the transparency ratings participants gave to the same meanings for those idioms during Session 1. As Table 2 shows, in Session 1, the transparency ratings for the learned and not-learned meanings were very similar. After the learning phase of Session 2, the mean for the learned meaning increased while the mean for the not-learned meaning decreased. The increase for learned items was significant [ $t(18) = 9.79, p < .0001$ ], as was the decrease for not-learned items [ $t(18) = 5.33, p < .0001$ ]. These results therefore replicate Keysar and Bly's (1995) findings: Meanings that are learned and used for idioms are perceived as more transparent than meanings that are not learned. Furthermore, learning one meaning actually depresses the perceived transparency of a different meaning for the same idiom.

The learning effect was similar for both true and alternative meanings. The overall ANOVA showed a main effect of meaning [ $F(1,17) = 22.11, p < .001$ ], reflecting somewhat higher ratings for true meanings over alternatives, but meaning did not interact with session [ $F(1,17) = 1.13$ ] or with learning [ $F(1,17) = 0.25$ ]. Looking at Session 2 alone, a 2 (meaning: true or alternative)  $\times$  2 (learning: learned or not learned) ANOVA showed a main effect of learning [ $F(1,18) = 103.89, p < .0001$ ], and no interaction of meaning with learning [ $F(1,18) = 1.55, n.s.$ ]. Looking across sessions at learned items only, the learned meanings of Session 2 were rated higher than the same meanings in Session 1, [ $F(1,18) = 95.87, p < .001$ ]. This effect did interact with meaning due to a slightly larger increase in ratings for the alternative meanings [ $F(1,18) = 5.07, p < .05$ ], but it was significant for both true and alternative meanings [ $t(18) = -6.77, p < .0001$  and  $t(18) = -9.15, p < .0001$ , respectively]. These outcomes indicate that regardless of which meaning participants are exposed to—the true idiomatic meaning or an alternative—learning the meaning increases the sense of transparency for that meaning.

At the conclusion of Session 2, participants had been told that during the experiment, they might not have learned the true idiomatic meaning. They were asked to go back through their transparency ratings for the 18 idioms and circle the meaning for each that they felt was the true meaning. Responses to the 8 idioms that they had been exposed to in the learning portion were extracted from the full set of 18 and scored. If participants circled the meaning that they learned during the experiment, the response received a score of 1; if they circled the meaning that they did not learn, it received a 0. If their choices were random, the mean score would be about .50. If they were systematically choosing the true meaning, their score would also be about .50, since each person learned true meanings for half the idioms and alternative meanings for the other half. The mean score was .90, which was significantly greater than .50 [ $t(18) =$

11.0,  $p < .001$ ]. Thus even when participants were alerted to the fact that the meaning that they had practiced might not be the one that is actually conventional for the idiom, they found the meanings they learned to be the most convincing candidates for the actual meaning. Notably, in Session 2, transparency ratings remained somewhat higher for true meanings than for alternatives [for both learned and not-learned items, with a main effect of meaning,  $F(1,18) = 15.66, p = .001$ , and no interaction with learning,  $F(1,18) = 1.52, p > .10$ ], supporting the notion that in this task participants were responding to genuine perceived transparency of the meanings and not to experimental demand to choose the meanings that they had been given previously.

### GENERAL DISCUSSION

Just as Keysar and Bly (1995) found for native speakers, when our participant non-native speakers of English were exposed to a meaning for an idiom that they did not know previously, their transparency ratings increased for the learned meaning. In addition, ratings decreased for a different possible meaning. Thus, the learned meaning made more sense to them after learning than before, and an unlearned meaning made less sense. Because the use of non-native speakers permitted us to present common, current English idioms as the stimuli, the data indicate that Keysar and Bly's findings generalize beyond the obsolete idioms that they used and are not due to properties peculiar to the obsolete idioms, such as particularly weak motivations for the true meanings and less inherent transparency. The data also eliminate the possibility that Keysar and Bly's findings occurred as a result of some form of demand characteristic or response strategy resulting from the fact that their native speaker participants would have recognized that these expressions were not part of the English language as they knew it. In addition, our data from the final task of Session 2, in which participants were explicitly asked to try to identify the correct meaning of each idiom, argue against the possibility that the effect on transparency ratings observed by Keysar and Bly and by us is due to participants' assuming that the meaning that has just been taught is the desired response. (The fact that ratings for the true meanings remained slightly higher than those for alternatives in Session 2, both when learned and when not learned, also indicates that participants were truly responding to a sense of transparency rather than to a guess about experimenter expectations.) These results support the hypothesis that the sense of transparency for meaning of an idiom is due, at least in part, to explanations generated after learning rather than from an a priori connection between the words of the expression or conceptual structures and the meaning. Further, they support the notion that once an explanation has been generated for one meaning, explanations that could motivate alternative meanings become less accessible.

The idea that language users actively construct explanations that link the words of an idiom to the meaning that they learn for it is consistent with observations out-

side the domain of figurative language. Humans in general have a strong inclination to generate causal explanations for phenomena (e.g., Gopnik, 2000), and in particular to generate explanations that link one observation to another (e.g., linking exposure to agents in the environment with illnesses; Kim & Keil, 2003). Perhaps most closely related to the phenomena of this study, Fischhoff and colleagues have shown that once people have acquired a piece of knowledge, they tend to feel that it is obvious that this fact would be true, although they could not have guessed that it would be true in advance (see, e.g., Fischhoff, 1982). For instance, when told about a scientific experiment and an outcome, people tend to see that outcome as inevitable and hence not surprising (Slovic & Fischhoff, 1977). However, if told a different outcome for the same experiment, they are equally sure that that outcome is the inevitable, and hence not surprising, one. In addition, they tend to think that someone else who has no knowledge of the outcome will be able to predict it—but the outcome that they think that person will predict is whichever one they themselves have learned as true. Fischhoff and colleagues argue that this "hindsight bias" comes about because once told an outcome, people attempt to make sense of the information by generating reasons why that outcome was obtained. In doing so, they deemphasize information that seems irrelevant in light of the outcome, which reduces the impression that any alternative could have come about. The sense of transparency that people acquire for the meaning of idioms, and the consequently reduced transparency of alternatives, may be a reflection of the general tendency of humans to engage in causal reasoning to explain observations and the reduced accessibility of alternative explanations that results when an explanation is generated. (See Keysar, 1994, for related phenomena in which knowing one interpretation of an utterance seems to block constructing other interpretations.)<sup>1</sup>

Does this sort of post facto generation of explanations fully explain the sense of transparency that language users have for idiomatic expressions, or does some of it, indeed, derive from the literal meanings of the words or from invoking conceptual metaphors in the comprehension process? We do not wish to argue that an idiom could have any meaning at all and still be transparent. Clearly, the fact that meanings that are the converse of an actual idiom meaning can be viewed as being just as transparent suggests that the individual words or the existence of particular conceptual metaphors do not provide strong constraints on what an idiomatic phrase can mean and what meaning can seem transparent. However, this observation is still compatible with the idea that conventional word meanings and conceptual metaphors do place some constraints on what meanings can be assigned and be transparent. Could *to spill the beans* transparently mean "to dislike one's in-laws" or "to play a great game of tennis"? The noun "beans" and the verb "spill" carry conventional meanings that in some way must be mapped to the idiomatic meaning in order for it to be transparent, and there may need to be some existing conceptual metaphor to facilitate or underlie the mapping. Although

surely it is possible for *to spill the beans* to have either of these meanings, just as *to kick the bucket* can mean “to die,” the meanings will not necessarily be transparent. Indeed, if there were no constraints on what meanings can be transparent for an idiom, then all idiomatic meanings should be transparent—but some are not.

The possibility that the sense of transparency for an idiomatic meaning arises in part from explanations generated after learning, does, however, highlight the need to distinguish between explanations of how idioms come to have a particular conventional meaning for a linguistic community and how individual language users process idioms in production and comprehension. On the one hand, intuitions about the connection between an idiom and its meaning cannot be taken as definitive evidence of the original motivation for an idiom. It is known that meanings of individual words sometimes shift over time to the point where a word means virtually the opposite of a previous meaning (e.g., a few centuries ago, *awful* meant “worthy of awe”; now it means “terrible”; *OED*, 1989). Idiomatic meanings may also evolve over time, perhaps even to the point where meanings reverse themselves. A case in point is *to break the bank*. In everyday speech, this expression generally means using up all of one’s money or becoming bankrupt. However, in the context of game shows and gambling, it means to win all the money at stake and increase one’s wealth (and, in fact, the *OED* [1989] lists both meanings). The second meaning probably derived from the first; if one breaks the game owner’s bank, one wins all the money. Only historical linguistic research can reveal the original motivation behind an idiomatic expression and how it may have shifted over time.

On the other hand, once an individual language user has learned a conventional meaning for an idiom and generated an explanation linking it to the words of the idiom, the links that are formed may have consequences for the mental operations that take place in comprehension when the idiom is encountered, or in production in the course of mapping thoughts to be expressed into words. Those linking explanations may be invoked either as an essential part of the comprehension or production process itself, or as a by-product of it. The fact that they may have been acquired as a result of learning the conventional meaning rather than in the original grasping of the meaning does not eliminate their potential processing relevance, just as once one learns the outcome of a scientific experiment and generates an explanation, that explanation is likely to pervade subsequent thinking about the experiment. Although observations of processing effects cannot constitute evidence about the original motivation for the idiom, they may be important to understanding idiom use.

## REFERENCES

- BOBROW, S. A., & BELL, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, *1*, 343-346.
- BORTFELD, H. (2003). Comprehending idioms cross-linguistically. *Experimental Psychology*, *50*, 217-230.
- CACCIARI, C., & GLUCKSBERG, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 217-240). Amsterdam: North-Holland.
- ENGEL, M., & GLUCKSBERG, S. (2003). *Idioms and ethnicity*. Manuscript in preparation.
- FISCHHOFF, B. (1982). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 335-351). Cambridge: Cambridge University Press.
- GLUCKSBERG, S. (1991). Beyond literal meanings: The psychology of allusion. *Psychological Science*, *2*, 146-152.
- GOPNIK, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory formation system. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 299-323). Cambridge, MA: MIT Press.
- IRUJO, S. (1986). Don’t put your leg in your mouth: Transfer in the acquisition of idioms in a second language. *TESOL Quarterly*, *20*, 287-304.
- KEYSAR, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, *26*, 165-208.
- KEYSAR, B., & BLY, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory & Language*, *34*, 89-109.
- KEYSAR, B., & BLY, B. (1999). Swimming against the current: Do idioms reflect conceptual structure? *Journal of Pragmatics*, *31*, 1559-1578.
- KIM, N. S., & KEIL, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & Cognition*, *31*, 155-165.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LAKOFF, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
- Longman dictionary of English idioms* (1979). Harlow, U.K.: Longman.
- NUNBERG, G., SAG, I. A., & WASOW, T. (1994). Idioms. *Language*, *70*, 491-538.
- Oxford English dictionary on-line* (2ND ED.) (1989). Oxford: Oxford University Press.
- SLOVIC, P., & FISCHHOFF, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception & Performance*, *3*, 544-551.
- SWINNEY, D., & CUTLER, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning & Verbal Behavior*, *18*, 523-534.

## NOTE

1. At first glance, the idea that people have a natural inclination to generate explanations so that idioms make sense to them may seem inconsistent with the difficulty that second language learners experience learning idioms of the second language. But the observed difficulty may, in fact, be quite consistent with the explanation-generation notion. Many idioms are language specific (e.g., Bortfeld, 2003; Irujo, 1986), and so there is no help in the acquisition of their meaning from the native language—unlike, for instance, the case of common nouns, where knowing the meaning of *chair* in English bootstraps acquisition of the meaning of *chaise* in French and so on. Because the words of idioms underconstrain their interpretation, even the meanings of those later viewed as transparent must be somehow grasped as a (partially) arbitrary convention first. The subsequent generation of an explanation may help the learner link the expression and its meaning in memory and use it, but it will not help the learner initially deduce the meaning. (And although our experimental passages were designed to make a meaning clear from a single exposure, ordinary discourse may be less helpful.) Of course, idioms such as *shoot the breeze*, which even native speakers do not find transparent, must be learned and retained entirely by rote memorization. These may be the idioms that produce the greatest learning difficulty (Engel & Glucksberg, 2003).



**APPENDIX**  
**The Eighteen Idioms Used in Session 1**

---

1. Learn the ropes
  2. Read the riot act
  3. Run rings around\*
  4. Break the bank\*
  5. Hold the fort\*
  6. Throw down the gauntlet\*
  7. Get your goat
  8. Put out to pasture\*
  9. Grin and bear it\*
  10. Put hairs on his chest
  11. Make your hair stand on end\*
  12. Harm a hair on his head
  13. Live from hand to mouth
  14. Have her eating out of the palm of your hand\*
  15. Sit on your hands
  16. Talk through his hat
  17. Bury the hatchet
  18. Lay your head on the block
- 

\*Idioms used in Session 2.

(Manuscript received August 25, 2003;  
revision accepted for publication January 13, 2004.)