Missing information in multiple-cue probability learning

CHRIS M. WHITE and DEREK J. KOEHLER
University of Waterloo, Waterloo, Ontario, Canada

In a multiple-cue probability learning task, participants learned to use six discrete symptoms (i.e., cues) to diagnose which of three possible flu strains a hypothetical patient suffered from. For some patients, information regarding the status of certain symptoms was not available. Various possible ways in which the missing cue information might be processed were distinguished and tested in a series of three experiments ($Ns = 80, 109$, and 61). The results suggest that the judged probability of the outcome variable (i.e., flu strain) was assessed by "filling in" the missing cue information with a mean value based on previous observations. The predictions of other methods of processing missing cue information are inconsistent with the data.

We frequently make predictions on the basis of a set of probabilistic cues—for instance, which team will win the sports game tonight or whether it will rain today. Often, all of the relevant information is not available. For instance, when one is looking at the sky to predict whether or not it will rain, one probably does not know whether the barometric pressure is rising or falling, which otherwise would have been factored into one's judgment. It is not only important to understand how people process the information that is available in these types of tasks, but also to understand how people respond when potentially predictive information is *not* available.

Studies of multiple-cue probability learning typically use tasks with complete information availability, but situations in which one has partial information are far more common. In this article, we briefly review conclusions that others have drawn regarding the effects of some information's being unknown (which we also refer to as missing or partial information) on people's judgments and then present three experiments in which we investigated the issue further.

There are several ways in which missing information can be handled when one is evaluating the probability of an outcome or hypothesis. We use the evidential support accumulation model (ESAM; Koehler, White, & Grondin, 2003) of how people generate probability judgments in a multiple-cue probability learning task to derive predictions that follow from different possible methods of processing

missing information. ESAM was developed using the framework of support theory (Tversky & Koehler, 1994); thus, that theory, as well as the relevant aspects of ESAM, are reviewed before the predictions of the different methods for processing missing information are outlined. Three experiments involving a multiple-cue probability learning paradigm in which six symptoms probabilistically discriminate among three possible flu strains are then described that test the predictions of the alternative methods for processing missing information.

Previous Research

In some of the previous research that used tasks involving partial information, participants have been required to estimate the status of an unknown cue (e.g., Yamauchi & Markman, 2000). When we talk of partial information in this paper, we refer to a situation in which participants learn to use a certain set of diagnostic cues and are then asked to judge category membership or hypothesis likelihood on the basis of only a subset of those cues. Most previous research using this type of design has not focused on the differences between the judgments given in the partial information condition and those given when complete information is available (e.g., Cuqlock-Knopp & Birch, 1988; Medin, Altom, Edelson, & Freko, 1982).

Ganzach and Krantz (1990) proposed that cues with unknown values or status (hereafter referred to as "unknown cues," for brevity) are assumed to have their previously observed mean value or status. This accounted for people's judgments in a task quite different from the one investigated here. In one experiment, participants predicted a hypothetical student's academic achievement (measured by GPA) on the basis of measures of motivation and intelligence. They based their judgments on prior knowledge and so received no feedback. The dependent measure was how regressive to the mean the predictions were when based on only one of the two cues. Participants who had previously made judgments based

on both cues made more regressive judgments when only given one cue value than did participants who had previously only made judgments based on one of the cues. In a subsequent experiment, one group of participants was told that the second cue existed, but never had experience using it. The regressiveness of these participants' judgments was no different from that of the judgments of participants who heard no mention of the second cue. In a task involving more novel predictor and criterion variables (car velocity and tire wear as predictors of braking distance), feedback was given. The same difference in the regressiveness of the judgments based on one of the cues was observed for approximately the first quarter of the 240 trials. However, as the experiment proceeded, the participants exposed to only one cue began to make their judgments essentially as regressive as the judgments of the participants exposed to both cues.

Ganzach and Krantz (1990) interpreted these results as showing that when participants made predictions from two cues, they set up a model to combine the two values to generate a judgment. When asked to generate a prediction on the basis of only one cue, the participants used the same model as before and inferred that the other cue was at its mean value and therefore made their judgments reasonably regressive. This is in contrast to the participants who only had experience making predictions based on one of the cues, who simply matched the extremity of the prediction to the extremity of the known cue (Kahneman & Tversky, 1973); this caused their predictions to be less regressive than those of the former group.

Gluck and Bower (1988) presented partial and full information trials when eliciting probability judgments from their participants. Their task involved four symptoms that probabilistically predicted two diseases. In some trials, they told participants that one cue was present and gave no information about the other three cues (partial information trials). On all other trials, the participants knew the status of all four cues (full information trials). One can compare the probability judgments given in each partial information trial to the judgment given in the full information trial that had the same cue present but all others absent (instead of unknown). People gave less extreme judgments in the partial information trials (they were spread over a range of approximately 0.45) than in the matched full information trials (in which the judgments were spread over a range of 0.63). Gluck and Bower concluded that participants had not treated unknown symptoms as being equivalent to absent symptoms. Nosofsky, Kruschke, and McKinley (1992) found similar results when substitutive cues were used (e.g., stuffy vs. runny nose) instead of present/absent cues. These researchers (Gluck & Bower, 1988; Nosofsky et al.,1992) considered only one other method of processing unknown cues, and that was to ignore them when generating a response, which was then implemented in the adaptive neural network model developed by Gluck and Bower.

While previous research has offered some insights, the data collected to date do not distinguish among a number of possible ways in which people may process missing information when making probability judgments. In the present experiments, a far richer data set was obtained with which to evaluate different possible methods of processing unknown cues.

Overview of the Present Research

Precise predictions from different possible methods of processing unknown cues can only be derived within the context of a broader model of how probability judgments are generated. We use ESAM to derive such predictions within the context of a multiple-cue probability learning task.

We used a simulated medical diagnosis task in which the participants learned the relationship between cues (six present/absent symptoms) and hypotheses (three flu strains). Each cue's status was probabilistically related to each of the hypotheses. After attempting to diagnose a series of patients in order to learn the cue–hypothesis relationships, participants judged the probability of a designated flu strain given the status of all of the symptoms or only the status of some of the symptoms for a particular patient. Comparing the two sets of judgments allows us to assess the effects of not knowing the status of some symptoms on people's probability judgments. The results can then be evaluated in the context of ESAM to test the predictions of different possible methods of processing unknown cues.

A probability judgment was elicited for each possible hypothesis given a particular pattern of cues. One measure that can be extracted from these judgments that illuminates many of the underlying processes is the sum of the probability judgments assigned to the competing hypotheses given a particular pattern of cues. Although probability theory requires that these probabilities sum to 100%, in previous research involving this type of task, when there are more than two hypotheses the probability judgments typically sum to more than 100% (Koehler, 2000; Koehler et al., 2003).

Support Theory

Support theory (Tversky & Koehler, 1994) accounts for the counterintuitive phenomenon of total probability judgments exceeding 100%, a finding that has been observed in multiple domains. The theory asserts that probability judgments are generated by comparing the support for each hypothesis, and that when multiple hypotheses are packed together into one conglomerate, they lose support so that the support for the conglomerate as a whole is less than the sum of that for its individual component hypotheses. The relationship between the support for the conglomerate and that of its components is said to be *subadditive*.

When a probability judgment is required and there are three or more hypotheses, the support for the hypothesis being judged (the *focal* hypothesis) must be compared with the support for multiple alternative hypotheses, which are packed together into a *residual* hypothesis and thereby lose support. This discounting of support for the

residual hypothesis causes the probability judgment for the focal hypothesis to be positively biased. When a probability judgment is elicited for each hypothesis, the residual is discounted each time and so each probability judgment is positively biased, resulting in the total probability judgment for all hypotheses for a given cue pattern being greater than 100%.

Koehler, Brenner, and Tversky (1997) proposed an extension of this theory. They posited that as the support for the focal hypothesis increases, the residual hypothesis is discounted more heavily. According to this principle of *enhanced residual discounting*, as the amount of evidence for the focal hypothesis increases, the evidence in support of the alternatives is less fully considered in judgments of probability.

ESAM

Support theory was developed to account for judgments in a wide range of tasks, and so did not state how the support for each hypothesis is actually computed. Koehler et al. (2003) developed ESAM as a model of how support is assessed for each hypothesis in a multiple-cue probability learning environment. The model compares the resulting support values using the framework of support theory and the assumption of enhanced residual discounting. For a full description of ESAM, see Koehler et al. (2003). Quantitative data fitting is unnecessary to test the predictions of the methods of processing unknown cues considered here, and so only the qualitative predictions of the model need to be described.

Of interest in the present research is that the total probability judgments exhibit systematic variability beyond simply exceeding 100%, which can be accounted for by ESAM. We use dichotomous cues (symptoms) that are either present or absent (e.g., whether or not the patient has a sore throat). In this type of task, previous studies have shown that the cue patterns that include a larger number of present symptoms receive higher total probability judgments than those with fewer present symptoms (Koehler, 2000; Koehler et al., 2003).

ESAM captures this effect because of its two main assumptions. First, ESAM assesses the evidential support for each hypothesis by *summing* the diagnostic value of the status of each cue. Second, to capture the observation that a cue's presence is more salient than is its absence (as in judgments of causation and covariation, e.g., Kao & Wasserman, 1993; Smedslund, 1963), a symptom's presence is assumed to contribute a larger amount to the support for each hypothesis than does its absence. These two assumptions result in cue patterns with more present cues conveying higher support values on average than cue patterns with fewer present cues. Following the principle of enhanced residual discounting, ESAM therefore implies that when more cues are present in a cue pattern, more residual discounting occurs because of the overall increased support for the focal hypothesis. More discounting of the residual when making each judgment results in a higher total probability judgment for cue patterns with

more present symptoms than for those with fewer present symptoms, consistent with the observed data.

Therefore, the greater salience of present symptoms and greater residual discounting when there is more support for the focal hypothesis combine to account for the total probability judgments increasing as the number of present cues in a pattern increases. We frame our analysis of the effects of how unknown cues are processed around this account.

Unknown Cues in ESAM

Although ESAM can account for the effects of the number of cues that are present or absent in a cue pattern on the total probability judgments, no provision is made in ESAM to account for the effect of unknown cues. We can assess how unknown cues are processed by determining what contribution they make to the support for each hypothesis in contrast to the contribution that the same cues give when they are present or absent. We do this by looking at the total probability judgments for patterns in which some cues are unknown and compare those with the total probability judgments for patterns in which those same cues are present, and in which those same cues are absent.

Unknown cues could affect the support assessment process in a variety of ways. Below, alternative methods for processing unknown cues are outlined and the pattern of total probability judgments that they predict is described. Figure 1 shows the predictions of the different methods graphically. In Figure 1 and throughout this paper the term *critical cues* refers to the cues that are known (i.e., to be present or absent) in the full information judgments but unknown in the partial information judgments, and the term *noncritical cues* refers to those cues that are known in both the full and partial information judgments.

The simplest method of processing unknown cues is to ignore them. Unknown cues would therefore contribute nothing to the total support for each hypothesis. We call this method *ignore*. When implemented within ESAM, it predicts that total probability judgments for partial information patterns will always be equal to (if absent cues are also given zero weight in the support assessment process) or less than (if absent cues are given some nonzero weight) the total probability judgments for the corresponding full information patterns in which the critical cues are absent.

The next simplest method is to assume that unknown cues are assigned a default status. If the cue is a binary symptom (present or absent, as in our studies), then the critical cues could be assumed to be absent (*assume-absent* method), and so the diagnostic value given by the unknown cue being absent is added for each unknown cue. This method predicts that total probability judgments for partial information patterns will be equal to the total probability judgments for corresponding full information patterns in which the critical cues are absent. Alternatively, one can assume that the critical cues are present (*assume-present* method). One would therefore
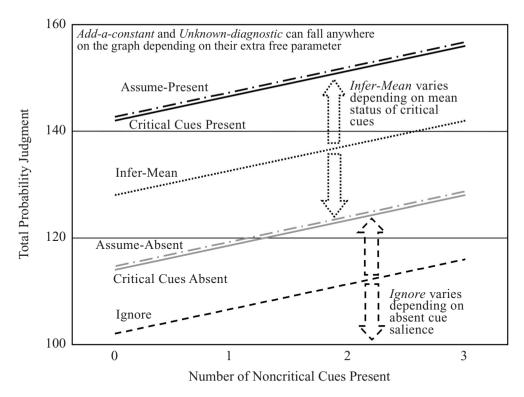
Figure 1. Total probability judgments for cue patterns containing certain numbers of noncritical cues present given certain statuses of the critical cues predicted by the different methods of processing unknown cues.

use the diagnostic value given by the unknown cue being present and so predict that total probability judgments for partial information patterns will be equal to the total probability judgments for corresponding full information patterns in which the critical cues are present.

Alternatively, one can infer that unknown cues are at their previously observed mean status (*infer-mean* method). In the case of binary (present/absent) cues, this amounts to setting the diagnostic value associated with the unknown cue equal to a weighted average of the diagnostic value for the cue's presence and for its absence, where the weight reflects the relative frequency with which the cue has been observed to be present versus absent. Effectively, this method adds the typical level of support conveyed by a given cue when it is known, and predicts that the total probability judgments for partial information patterns will fall somewhere between the total probability judgments given for corresponding patterns in which the critical cues are absent and in which they are present. Exactly where they fall between these two extremes is determined by the previously observed mean status of the critical cues and the response-scaling free parameter that is already part of ESAM.

Another possibility would be to add a constant amount (which does not vary across cues) to the evidential support for each cue whose status is unknown (*add-a-constant* method). This proposal is based on the intuition that in

the face of the uncertainty produced by an unknown cue, the individual might "hedge" his/her assessment by adding a fixed amount of support to each hypothesis in response to each unknown cue. This approach may be simpler than the infer-mean method. The amount of support added for each unknown cue would have to be estimated by a free parameter, which would determine whether the total probability judgments for partial information patterns would be greater than, equal to, or less than the total probability judgments for full information judgments.

A final method assumes that the cue being unknown is itself diagnostic information, just as is its being present or absent. We call this the *unknown-diagnostic* method. As stated above, we assume that the presence and absence of a cue have different saliencies, and so ESAM differentially weights their respective diagnostic values when assessing the support for each hypothesis by use of a free parameter. In a similar way, the relative salience of unknown cues would need to be reflected in the weight given to the diagnostic value of an unknown cue by use of a free parameter (with the diagnostic value of a cue's unknown status being computed in the same way as is the diagnostic value of a cue being present or absent). This method can therefore predict the total probability judgments to be higher or lower for partial information patterns than for full information patterns, depending on the value of the free parameter (which could be manipulated by varying

the salience of unknown cues). This method reduces to the add-a-constant method when a cue's being unknown is not diagnostic, as in our experiments.

We now describe three experiments designed to investigate which of the models based on various possible methods of processing unknown cues best captures the observed data. The first two experiments were part of a broader investigation by White (2002). They were designed to not only study the effects of missing information on probability judgments but also to investigate how sensitive people's judgments are to intercue dependencies (i.e., when the predictive value of one cue depends upon the status of another cue). The results regarding sensitivity to the intercue dependencies are reported elsewhere (White & Koehler, 2003). Here, we focus only on the results pertaining to the processing of the unknown cues. The third experiment was conducted later for the sole purpose of clarifying how people process unknown cues.

EXPERIMENT 1

We used a multiple-cue probability learning paradigm with six cues (A–F) and three hypotheses (Flu Strains 1, 2, and 3). There were two cue structures. In the independent-diagnostic cue structure, all six cues were conditionally independent of each other and each gave diagnostic information regarding the likelihood of the three flu strains. In the dependent-nondiagnostic cue structure, the cue–hypothesis co-occurrence frequencies were the same as in the independent-diagnostic cue structure, but cues D–F were redundant with and imperfectly dependent upon cues A–C. This was achieved by grouping the six cues into pairs (cues A and D were paired, as were cues B and E, and cues C and F); the first cue in each pair (cues A–C) gave diagnostic information and the second cue in the pair (cues D–F) depended only on the status of cues A–C and not directly on the patient's flu strain. Cues D–F therefore gave no information beyond that which cues A–C gave regarding the likelihood of the three flu strains. The differences between the cue structures is the focus of our other report on this experiment (White & Koehler, 2003). Of importance to the present research is that after being exposed to all six cues on every trial of the training phase, in one block of the subsequent judgment phase participants only had information about cues A–C, while in the other block they had information concerning all six cues.

Method

Participants

Eighty volunteer participants were recruited from introductory psychology courses at the University of Waterloo and participated for course credit.

Procedure

Participants completed the task in separate rooms. The task was administered and all instructions were displayed by a computer.

Training phase. Participants saw 252 fictitious patients in the training phase who each suffered from one of three flu strains (named Flu Strain 1, 2, and 3) in a random order. On each trial, a computer presented the participant with a patient who had a fever plus a certain combination of the six symptoms (named headache, sore throat, earache, chills, dizziness, and cough). Each patient either suffered from a symptom (denoted by the name of the symptom appearing in capital letters, e.g., HEADACHE, which we refer to as a "present" symptom) or did not suffer from the symptom (denoted by the name of the symptom appearing in lowercase letters with the word *no* in front of it, e.g., "no headache," which we refer to as an "absent" symptom). Each symptom was probabilistically related to the three flu strains (referred to as hypotheses here), but the participants had no knowledge of what the relationships were at the beginning of the training phase. Participants selected which of the three flu strains they believed that each patient was suffering from on the basis of the patient's pattern of symptoms. This was immediately followed by feedback indicating whether their choice was correct and which flu strain the patient actually had.

For each participant, different symptom names were randomly assigned to the six symptoms. All of the symptoms were displayed on the screen in a constant spatial order for each participant, but the spatial order varied randomly between participants.

Judgment phase. Following completion of the training phase, participants read instructions regarding the judgment phase. In these trials, participants saw a fictitious patient who had a certain combination of symptoms (plus fever) and estimated the probability that the patient was suffering from a designated flu strain (by selecting a probability judgment between 0% and 100% in 10% increments). In this phase, participants were not told which flu strain the patient actually suffered from.

The judgment phase consisted of two blocks. In the full information block, there were 192 trials in a random order in which the participants were provided with information about all six symptoms. In the partial information block, there were 24 trials in a random order in which the participants only had information about symptoms A, B, and C; symptoms D, E, and F (the unknown symptoms) were not shown on the screen. Participants were informed that if a symptom was not shown, then this gave no information about its status, and could be either present *or* absent. The participants completed the two blocks of judgments in a counterbalanced order. The order of the two judgment blocks did not affect the results of the analyses, and so the data were collapsed across the two orders.

Materials

Training phase. Table 1 shows the number of times that each symptom was present and absent in the training phase when a patient had a certain flu strain. The prior probabilities or "base rates" of the three flu strains were equal. The frequency counts shown in Table 1 were the same in both cue structures; only the frequency with which certain *pairs* of cues occurred with each hypothesis differed. The results from the two cue structures did not differ in regard to the analyses presented below, indicating that our results generalize across different types of cue structures. For ease of presentation,

Table 1
Cue Frequency Counts for the Training Phase of Experiment 1

Symptom	Flu Strain 1 (n = 84) +	Flu Strain 1 (n = 84) −	Flu Strain 2 (n = 84) +	Flu Strain 2 (n = 84) −	Flu Strain 3 (n = 84) +	Flu Strain 3 (n = 84) −	Total (n = 252) +	Total (n = 252) −
A	76	8	25	59	25	59	126	126
B	25	59	76	8	25	59	126	126
C	25	59	25	59	76	8	126	126
D	66	18	30	54	30	54	126	126
E	30	54	66	18	30	54	126	126
F	30	54	30	54	66	18	126	126

Note—+, cue is present; −, cue is absent.

the data are collapsed across the two cue structures in the results reported below.

Judgment phase. In the full information block, the 192 trials consisted of each of the 64 possible six-symptom patterns combined with each of the three flu strains. In the partial information block, the 24 trials consisted of each of the eight possible three-symptom patterns combined with each of the three flu strains.

Results

A data collection problem resulted in the deletion of the data from 1 participant. The accuracy (i.e., proportion of trials on which the correct diagnosis was chosen) of the remaining participants during the second half of the training phase was 58%. This compares with chance accuracy of 33% and with the maximum theoretical accuracy of 79% and 68% in the independent-diagnostic and dependent-nondiagnostic cue structures, respectively. The maximum theoretical accuracy is defined as that achievable by someone who knows all of the cue–hypothesis relationships perfectly before the beginning of the training phase, and uses them in a Bayesian manner to select the most likely hypothesis on each trial. Participants appeared to have learned the cue–hypothesis relationships reasonably well, although not perfectly.

Since the status of cues D–F was not known on the partial information trials, we refer to these as the *critical cues*; the status of cues A–C was always known and so we refer to these as the *noncritical cues*. The data from the

pattern in which all of the cues (critical and noncritical) were "absent" and the pattern in which all of the cues were "present" were excluded from the analyses below because prior research suggests that probabilities are judged differently for these special cases than they are for the more typical patterns (Koehler, 2000; Koehler et al., 2003).

Our primary dependent measure is the total probability assigned to the three possible flu strains given a particular cue (i.e., symptom) pattern. Normatively, this value should always be 1, or 100%. Figure 2 shows the mean total probability judgment as a function of the number of present noncritical cues, separately for patterns in which the critical cues were unknown, all present, or all absent. Consistent with previous research (Koehler, 2000; Koehler et al., 2003), the mean total probability judgments are consistently greater than 100% (i.e., they display subadditivity) and increase with the number of present cues.

We conducted an analysis of variance (ANOVA) to assess the effects of the status of the critical cues (present, absent, or unknown) and the number of noncritical cues present (one or two). The status of the critical cues affected the total probability judgments [$F(2,156) = 38.5$, $MS_e = 688.5$, $p < .001$], as did the number of noncritical cues present [$F(1,78) = 43.9$, $MS_e = 601.6$, $p < .001$]. These two factors did not interact ($F < 1$). More specifically, when the critical cues were unknown, the total probability judgments were different from those given when the critical cues were absent [$F(1,78) =$
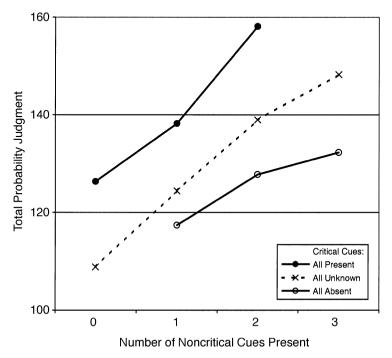

Figure 2. Total probability judgments for cue patterns containing certain numbers of noncritical cues present given certain statuses of the critical cues in Experiment 1.

16.6, MS_e = 927.1, $p < .001$] or present [$F(1,78)$ = 30.0, MS_e = 1,114.9, $p < .001$].

When the critical cues were unknown, the total probability judgments fell between those given when the critical cues were all present and all absent. In testing the predictions of various possible methods of processing unknown cues, it is helpful to have a simple measure of where in this range the total probability judgments fell when the critical cues were unknown. To measure this, we compare the difference between the mean total probability judgment given when the critical cues were unknown (T_{unk}) and absent (T_{abs}) with the difference when the critical cues were present (T_{pres}) and absent (T_{abs}) to obtain a ratio R_{unk}:

$$R_{unk} = \frac{T_{unk} - T_{abs}}{T_{pres} - T_{abs}}. \qquad (1)$$

The value of R_{unk} can be compared with the value predicted by each of the methods of processing unknown cues described above. If unknown cues are ignored, then $T_{unk} \leq T_{abs}$, and so $R_{unk} \leq 0$. The assume-absent method predicts that $T_{unk} = T_{abs}$, and so $R_{unk} = 0$; the assume-present method predicts that $T_{unk} = T_{pres}$, and so $R_{unk} = 1$. The infer-mean method predicts that $T_{abs} \leq T_{unk} \leq T_{pres}$, and so $0 \leq R_{unk} \leq 1$. The add-a-constant and unknown-diagnostic methods predict that T_{unk} (and therefore R_{unk}) will vary depending on each method's extra free parameter.

The value of R_{unk} can be calculated for each participant. However, the range between the mean total probability judgment given when the critical cues were present and absent ($T_{pres} - T_{abs}$) is extremely small for some participants, making the denominator of R_{unk} very small, thereby making R_{unk} extremely large. To avoid the disproportionate impact of these extreme values, we removed the data from the participants whose range for the denominator was less than 4% on the probability scale. This value is somewhat arbitrary but was chosen in order to exclude all of the extreme values of R_{unk} in the three experiments reported here. Some non-extreme values of R_{unk} are also excluded by using this cutoff, but these are usually unreliable values anyway, as a very small change in T_{unk} would cause R_{unk} to change dramatically in these cases. In the present experiment, use of this cutoff value results in exclusion of the data from 2 participants. The mean value was R_{unk} = .40 (median = .33, SEM = .11), which was significantly different from 0 [$t(77)$ = 3.74, $p < .001$] and 1 [$t(77)$ = 5.70, $p < .001$]. The distribution of R_{unk} is unimodal, which argues against the possibility that the observed mean arises from a mixture of strategies (with some of the people using an assume-present strategy and some using an assume-absent strategy).

Before discussing the implications of this pattern of results, we first describe the results of a second study designed to establish the generalizability of our findings.

EXPERIMENT 2

Experiment 2 was similar to Experiment 1, except as follows. To ensure that the differences observed between the partial and full information judgments were not due to the two types of judgments being elicited in different blocks, the partial and full information trials were randomly mixed. In addition, to eliminate any possible novelty effects of the partial information trials in the judgment phase, partial information trials were included in the training phase. Instead of always having the status of cues D–F unknown in the partial information trials (i.e., designating cues D–F as the critical cues), all possible combinations of which three cues were known and which three were unknown were sampled (e.g., on one trial, only information about cues C, D, and F was given, and on another, only information about cues A, C, and E). The training phase was also lengthened to ensure that participants had ample opportunity to learn the cue structure.

Three cue structures were used, one in which cues D–F were imperfectly dependent upon cues A–C (dependent-nondiagnostic) and two in which all six cues were conditionally independent of each other (independent-diagnostic and independent-nondiagnostic). The cue structure again had no effect on the results of the analyses presented below, and so the data are collapsed across cue structures.

Method

Participants

One hundred and nine volunteer participants were recruited from introductory psychology courses at the University of Waterloo and participated for course credit.

Procedure

Training phase. The number of training trials was increased from 252 to 336, of which 240 were full information training trials and 96 were partial information trials. Training trials were again identical for all participants assigned to a given cue structure, but were presented in a different random order for each participant that mixed the partial information and full information trials.

Judgment phase. This phase consisted of 120 full information trials and 88 partial information trials randomly mixed together.

Materials

Training phase. See Table 2 for the frequencies that each cue had each status in conjunction with each flu strain over the course of the training trials.[1]

Table 2
Cue Frequency Counts for the Training Phase of Experiment 2

Symptom	Flu Strain 1 (n = 112)			Flu Strain 2 (n = 112)			Flu Strain 3 (n = 112)			Total (n = 336)		
	+	−	?	+	−	?	+	−	?	+	−	?
				All Cue Structures								
A	88	8	16	28	68	16	28	68	16	144	144	48
B	28	68	16	88	8	16	28	68	16	144	144	48
C	28	68	16	28	68	16	88	8	16	144	144	48
			Independent-Diagnostic and Dependent-Nondiagnostic Cue Structures									
D	76	20	16	34	62	16	34	62	16	144	144	48
E	34	62	16	76	20	16	34	62	16	144	144	48
F	34	62	16	34	62	16	76	20	16	144	144	48
				Independent-Nondiagnostic Cue Structure								
D	48	48	16	48	48	16	48	48	16	144	144	48
E	48	48	16	48	48	16	48	48	16	144	144	48
F	48	48	16	48	48	16	48	48	16	144	144	48

Note—+, cue is present; −, cue is absent; ?, cue is unknown.

Judgment phase. There are 192 possible full information judgments. However, only 40 of the 192 judgments are mathematically distinct. This is because the relationship that each cue within each set of three cues (i.e., A–C and D–F) had with each hypothesis is a mirror image of that had by one of the other cues in that set with one of the other hypotheses. For example, if we denote a present symptom with a capital letter and an absent symptom with a lowercase letter, then the probability judgment of Flu Strain 1 for cue pattern Abc def is logically equivalent to the probability judgment for Flu Strain 2 for cue pattern aBc def. To obtain a reliable measure of each of these 40 judgments, we elicited three judgments for each, yielding 120 full information judgments. Applying the same logic to the partial information judgments, we elicited one example of each of the 88 mathematically distinct judgments (out of the full set of 480 possible partial information judgments that have three cues known).

Results

A data collection problem resulted in the removal of the data from 1 participant. The accuracy of the remaining participants had asymptoted by the end of the training phase.[2] In the second half of the training phase, mean accuracy was 58%, 56%, and 51% for the independent-diagnostic, dependent-nondiagnostic, and independent-nondiagnostic cue structures, respectively, compared with chance accuracy of 33% and the maximum theoretical accuracy of 77%, 68%, and 66% in each cue structure, respectively. Again, participants appeared to have learned the cue–hypothesis relationships reasonably well, although not perfectly.

In this experiment, the three critical cues for a given judgment could be any possible subset of the six cues. Therefore, each pattern in which the critical cues were unknown was matched to corresponding patterns in which the critical cues were all present or absent. Figure 3 displays the results. This experiment incorporated many methodological changes from Experiment 1, yet the same effects were observed. The same factors were included in the ANOVA as in Experiment 1: the status of the critical cues (present, absent, or unknown) and the number of noncritical cues present (one or two). The status of the critical cues significantly affected the total probability judgments [$F(2,214) = 74.4$, $MS_e = 637.9$, $p < .001$], as did the number of noncritical present cues [$F(1,107) = 27.9$, $MS_e = 332.4$, $p < .001$], and again, these factors did not interact ($F < 1$).

The more specific contrasts that were evident in Experiment 1 were again evident here. The total probability judgments given when the critical cues were unknown were different from those given when the critical cues were absent [$F(1,107) = 31.0$, $MS_e = 585.1$, $p < .001$] and present [$F(1,107) = 44.5$, $MS_e = 950.9$, $p < .001$]. Only the magnitude of this latter effect depended on the number of noncritical cues [$F(2,214) = 5.73$, $MS_e = 393.4$, $p < .01$], but this interaction was not significant in any of the equivalent comparisons in Experiment 1.

The mean value of R_{unk} was .52 (median = .39, $SEM = .081$), which differed significantly from 0 [$t(102) = 6.44$, $p < .001$] and 1 [$t(102) = 5.89$, $p < .001$]. As in Experiment 1, we removed the data from the participants whose range for the denominator of R_{unk} was less than 4% before this analysis, which was the case for 5 participants.

Discussion of Experiments 1 and 2

Total probability judgments given when the critical cues were unknown were *higher* than those given to the
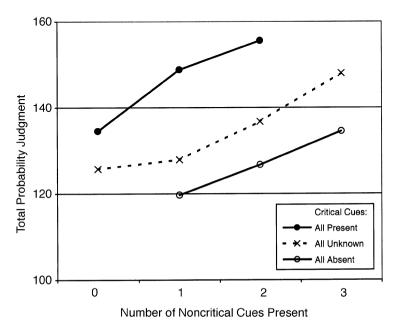

Figure 3. Total probability judgments for cue patterns containing certain numbers of noncritical cues present given certain statuses of the critical cues in Experiment 2.

full information judgments in which the critical cues were absent. This implies that people did not assume that the unknown cues were absent (the assume-absent method of processing unknown cues). This result agrees with the findings of Gluck and Bower (1988). It is not in line with the predictions of ignoring unknown cues (Figure 1), which predicts that the total probability judgments given when the critical cues were unknown would be less than or equal to those given to the full information judgments in which the critical cues were absent. In addition, the total probability judgments given when the critical cues were unknown were *lower* than those given when the critical cues were present, implying that people did not use the assume-present method. These effects appear to be robust across different cue structures. The cue structure did influence other aspects of the judgments, as reported elsewhere (White & Koehler, 2003), indicating that the consistent effects of unknown information are not attributable to a general insensitivity to differences between the cue structures.

Three of the methods for processing unknown cues described above can account for the observed results: the infer-mean, add-a-constant, and unknown-diagnostic methods. The proposal that unknown cues were inferred to be at their previously observed mean status (the infer-mean method) is arguably the most parsimonious of the three because it requires no additional free parameters to account for the observed results. However, further manipulations are needed to tease apart the predictions made by the three remaining methods.

EXPERIMENT 3

In Experiments 1 and 2, each cue was present half of the time and absent half of the time. In this experiment, we vary the frequency of cue presence versus absence to test a specific prediction of the infer-mean hypothesis—namely, that the total probability judgments given when the critical cues are unknown will vary on the basis of the mean status of the critical cues. The infer-mean method predicts this because it assumes that the diagnostic value associated with an unknown cue is equal to a weighted average of the diagnostic value for the cue's presence and for its absence, where the weight reflects the relative frequency with which the cue has been observed to be present versus absent.

In the training phase, three of the six cues (cues A, B, and C) were present 49% of the time and absent 51% of the time (high present-frequency cues), and the other three cues (cues D, E, and F) were present 24% of the time and absent 76% of the time (low present-frequency cues). Of the three methods of processing unknown cues consistent with the results of Experiments 1 and 2, only the infer-mean method predicts any difference in the partial information total probability judgments, depending on whether the critical cues are high present-frequency or low present-frequency cues. Specifically, the infer-mean method predicts that the value of R_{unk} will be higher

when the critical cues are high present-frequency cues than when they are low present-frequency cues. The two other methods of processing unknown cues that can also account for the data from Experiments 1 and 2 (the add-a-constant and unknown-diagnostic methods) predict this value to be constant across this manipulation.

We also varied the number of cues that were not known in the partial information pattern, with some participants being told the status of five of the six cues (the one-unknown condition) and the others being told the status of three of the cues (the three-unknown condition). The number of unknown cues was varied to generalize the results of Experiments 1 and 2 beyond the situation in which exactly half of the cues are unknown.

Method

Participants

Sixty-one volunteer participants were recruited from introductory psychology courses at the University of Waterloo and participated for course credit. They were randomly assigned to the two judgment phase conditions (31 to the one-unknown condition and 30 to the three-unknown condition).

Procedure

Training phase. The training phase consisted of 288 full information trials in a random order. Experiment 2 showed that the differences between the probability judgments observed in Experiment 1 were not caused by the novelty of the partial information trials in the judgment phase, and so partial information trials were not included in the training phase of this experiment.

Judgment phase. In the one-unknown condition, this phase consisted of 144 trials: 40 full information trials and 104 partial information trials. In the three-unknown condition, this phase consisted of 192 trials: 120 full information trials and 72 partial information trials. Full and partial information trials were again randomly mixed.

Materials

Training phase. Only one cue structure was used in which all of the symptoms were conditionally independent of each other. The number of times that each cue had each status in conjunction with each hypothesis is shown in Table 3.

To ensure that the relative frequency of the cue statuses was the only thing that differed between the two sets of cues, the diagnostic value of the cues for the three hypotheses was equated. Specifically, the information in the contingency table containing the three cues in each group, the two statuses of each cue, and the three hypotheses must be equated. Therefore, a method for measuring the strength of the relationship between multiple cues and multiple hypotheses in a $3 \times 2 \times 3$ multidimensional contingency table is re-

Table 3
Cue Frequency Counts for the Training Phase of Experiment 3

Symptom	Flu Strain 1 ($n = 96$)		Flu Strain 2 ($n = 96$)		Flu Strain 3 ($n = 96$)		Total ($n = 288$)	
	+	−	+	−	+	−	+	−
A	86	10	28	68	28	68	142	146
B	28	68	86	10	28	68	142	146
C	28	68	28	68	86	10	142	146
D	57	39	6	90	6	90	69	219
E	6	90	57	39	6	90	69	219
F	6	90	6	90	57	39	69	219

Note—+, cue is present; −, cue is absent.

quired. McGill (1954) developed such a measurement method, based on information theory, that measures the information transmitted in the form of the number of "bits" per signal. To correctly choose between two hypotheses (in this case, flu strains) one would need one bit of information, and to choose between four hypotheses, one would need two bits of information. Since this task involves three flu strains, between one and two bits of information would be needed to allow one to make the correct choice every time. By McGill's measure, cues A–C transmit 0.372 bits, and cues D–F transmit 0.373 bits per signal, suggesting that the diagnostic value of cues A–C and D–F were effectively equal.[3]

Judgment phase. In the one-unknown condition, we sampled each of the 40 distinct full information judgments once and each of the 104 distinct one-unknown partial information judgments once. In the three-unknown condition, the 120 full information trials consisted of three examples of each of the 40 mathematically distinct groups of judgments. Of the 88 distinct partial information judgments, we were most interested in those patterns in which either all three of the high present-frequency cues or all three of the low present-frequency cues were unknown. There are only 12 mathematically distinct judgments fitting these criteria. Again, each of these was sampled three times, yielding 36 more judgments. We randomly chose 12 of the remaining 76 distinct three-unknown partial information judgments and sampled each of these three times.

Results and Discussion

Accuracy had asymptoted by the end of the training phase. The mean for the second half of the training phase was 67%, compared with chance accuracy of 33% and the maximum theoretical accuracy of 82%. This suggests that the participants learned the cue–hypothesis relationships at least as well as did those in Experiments 1 and 2.

The data were split on the basis of whether the critical cues were low or high present-frequency cues. Since only one of the cues was unknown in each partial information pattern of the one-unknown condition, there was only one critical cue for each judgment in this condition. Whichever set of cues contained the critical cue was deemed to be the critical set of cues even though not all of the cues in the set were unknown. In the three-unknown condition, the only partial information judgments used for this analysis were those in which either all of the low or high present-frequency cues were unknown.

The predictions of the infer-mean method of processing unknown cues could be assessed by looking for interactions between the status of the cues (present, absent, unknown) and which set of cues were the critical cues (low or high present-frequency cues) in a way similar to how we assessed the effects in Experiments 1 and 2. The complete ANOVA includes three within-subjects variables: the status of the critical cues (absent, present, or unknown), which cues were the critical cues (low or high present-frequency cues), and the number of noncritical cues present (one or two). The ANOVA also includes one between-subjects variable: the number of unknown cues (one or three). The results of this ANOVA are very hard to interpret since the interaction between all four of these variables is significant [$F(2,118) = 4.57$, $MS_e = 1,038.9$, $p < .05$]. Therefore, even though the interaction between the two most important variables—whether the high or

low present-frequency cues are the critical cues, and the status of the critical cues—is significant [$F(2,118) = 3.56$, $MS_e = 1,038.9$, $p < .05$], the form of this interaction depends on the level of the other two variables. We therefore turn to the simpler and more direct predictions that the infer-mean method makes regarding the R_{unk} measure.

Two values of R_{unk} are computed for each individual, one for each of the critical cues being high and low present-frequency cues. Two variables (the number of cues present and the status of the critical cues) disappear in the analysis of R_{unk} because data from each level of these variables goes into this one aggregate measure (see the Results section of Experiment 1 for a complete description of how this measure is computed). The other variable that still remains is the number of unknown cues (one or three), which was varied between participants.

As predicted by the infer-mean method of processing unknown cues, the total probability judgments given when the critical cues were high present-frequency cues were higher than those given when the critical cues were low present-frequency cues (relative to when the critical cues were absent and present): median $R_{unk} = .52$ and .21, respectively [means: $R_{unk} = .82$ and .20, respectively; $F(1,52) = 5.43$, $MS_e = 1.89$, $p < .05$].[4] The number of missing cues had no effect on this measure and did not interact with whether the unknown cues were high or low present-frequency cues ($Fs < 1.0$). As in Experiments 1 and 2, we removed the data from the participants whose range for the denominator of R_{unk} was less than 4% for either the high or low present-frequency cues before these analyses, which was the case for 7 participants.

GENERAL DISCUSSION

The only method of processing unknown cues considered in this paper that can account for the results of Experiments 1–3 is the infer-mean method. This result appears to be quite robust, with the predictions of this method being confirmed across different cue structures (Experiments 1 and 2) and across differing numbers of unknown cues (Experiment 3). We therefore assert that once people have learned to use multiple cues to predict the likelihood of certain outcomes, if information concerning some of those cues is not available, they will infer that those cues are at their previously observed mean status and generate a judgment on the basis of that inference. It does not appear that people simply ignore those cues or give those cues a preset default value (absent or present), or treat the unknown status of each missing cue as diagnostic.

Even though we have framed the infer-mean method of processing unknown cues in terms of people "filling in" an unknown cue with its previously observed mean status on each trial and calculating a diagnostic value based on that inference, there are other process interpretations of this model. As we mentioned when introducing this model, the support given by an unknown cue

could be equated with the typical or average amount of support given by that cue on trials in which its status was known. This would yield identical predictions to those of the filling-in model, as would a third variant: People could infer that unknown cues are either present or absent on each individual trial, where the probability of making each inference reflects the relative frequency that the unknown cue was present and absent when it was known. The data from the present experiments cannot distinguish between these different psychological processes, all of which would have the same effect on the total probability judgments.

It could be argued that the methods of processing unknown cues proposed here would make different predictions in the framework of a model other than ESAM, and a different method of processing unknown cues might then be able to account for the present results. Without a fully developed model, however, specific predictions of the more complex methods of processing unknown cues are difficult to derive, so having a model is essential. There are two general types of models: those that assess information on a cue-by-cue basis and those that assess entire exemplars. We now address how well other models can account for the patterns of data that are of interest here, and how likely it is that the processes underlying these models will match the processes underlying people's behavior in this type of task.

In regard to other models that integrate information on a cue-by-cue basis (e.g., Gluck & Bower, 1988), although the rules governing how the outputs of these models are converted into probability judgments could be changed, currently none of the models is able to generate probability judgments that sum to more than 100% (subadditivity). In addition, even though other models explicitly allow for different cues to have varying saliencies, they do not explicitly allow for different cue statuses to have varying saliencies. One way around the latter problem would be to assume that a cue's presence and absence are represented separately. Each status of each cue could then have its own salience value. However, the separation at a representational level of a cue's presence and absence lacks empirical support. Even given these additional assumptions, it is not clear whether these other models would be able to account for the systematic variation in subadditivity as a function of the number of present cues.

Models that base their judgments on the similarity of the current exemplar to a set of stored exemplars (e.g., Medin & Schaffer, 1978) suffer from similar problems. In addition, many characteristics of the task used in the present experiments have been found by other researchers to encourage the use of cue abstraction and integration processes rather than exemplar-based processes (Juslin, Jones, Olsson, & Winman, 2003; Juslin, Olsson, & Olsson, 2003). Specifically, the task used here involved probabilistic rather than deterministic outcomes, propositional rather than pictorial cues, and the responses of interest (probability judgments) were more continuous

than categorical. Exemplar-based models are therefore based on processes that are arguably unlikely to be used by participants in this task. In addition, previous work in our lab using the same paradigm showed that when ESAM bases the support for each hypothesis on the similarity of the current exemplar to stored exemplars, the judgments produced fit the data less well than did the judgments produced when support was based on a cue-by-cue analysis (Koehler et al., 2003, pp. 182–184).

In any case, it is reassuring that our results agree with the findings of Ganzach and Krantz (1990), who also concluded that unknown cues are inferred to be at their mean value or status using a quite different approach from that used here. This converging evidence provides strong support for the hypothesis that if people are accustomed to basing their responses on a certain cue, and if the status or value of that cue is not available to them, they will continue to consider that cue in calculating their response, but will act as if the cue is at its mean status or value. If one must infer the value of an unknown cue, then using the previously observed mean value of that cue will generally produce reasonably accurate judgments. This behavior is adaptive if the mechanism involved in making estimates and judgments from imperfect cues forces one to choose a value for each relevant cue.

REFERENCES

Cuqlock-Knopp, V. G., & Birch, D. (1988). Cue selection and cue utilization as separate but coordinated processes in multiple-cue predictions. *American Journal of Psychology*, **101**, 555-560.

Ganzach, Y., & Krantz, D. H. (1990). The psychology of moderate prediction I: Experience with multiple determination. *Organizational Behavior & Human Decision Processes*, **47**, 177-204.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 227-247.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 924-941.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, **132**, 133-156.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237-251.

Kao, S.-F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1363-1386.

Koehler, D. J. (2000). Probability judgment in three-category classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 28-52.

Koehler, D. J., Brenner, L. A., & Tversky, A. (1997). The enhancement effect in probability judgment. *Journal of Behavioral Decision Making*, **10**, 293-313.

Koehler, D. J., White, C. M., & Grondin, R. (2003). An evidential support accumulation model of subjective probability. *Cognitive Psychology*, **46**, 152-197.

McGill, W. J. (1954). Multivariate information transmission. *Psychometrika*, **19**, 97-116.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 37-50.

MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

NOSOFSKY, R. M., KRUSCHKE, J. K., & MCKINLEY, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 211-233.

SMEDSLUND, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, **4**, 165-173.

TVERSKY, A., & KOEHLER, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, **101**, 547-567.

WHITE, C. M. (2002). *Modeling the influence of cue dependencies and missing information in multiple-cue probability learning.* Unpublished master's thesis, University of Waterloo.

WHITE, C. M., & KOEHLER, D. J. (2003). *Modeling the influence of cue dependencies in multiple-cue probability learning*. Unpublished manuscript, University of Waterloo.

YAMAUCHI, T., & MARKMAN, A. B. (2000). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory & Cognition*, **28**, 64-78.

NOTES

1. Every cue had an equal chance of being unknown in each partial information trial, and so each was unknown in 48 of these 96 trials. The fact that a cue was unknown had no diagnostic value because the number of times it was unknown in conjunction with each flu strain was the same (16 times each).

Two of the cue structures were similar to those used in Experiment 1. Cues A–C were always diagnostic in all three cue structures. In the independent-diagnostic cue structure, Cues D–F were also always diagnostic. In the dependent-nondiagnostic cue structure, Cues D–F were nondiagnostic when the status of Cues A–C was known and diagnostic when the status of Cues A–C was not known. For the sake of completeness, an independent-nondiagnostic cue structure was added where Cues D–F were always nondiagnostic.

2. This was determined by arbitrarily splitting the training phase into six blocks and assessing that the mean accuracy per block had leveled off by the last training block.

3. The cue structure used also yields similar results on several other measures of the strength of association. We describe one more method to illustrate this. The mathematically correct probabilities were calculated using the frequencies that each cue had each status in conjunction with each flu strain during the training phase. These frequencies were combined using Bayes's rule (assuming conditional independence) to generate one probability for each flu strain for each possible full information pattern. The resulting 192 probabilities were analyzed to measure how much the status of each set of cues affected them.

Each cue pattern was matched with a corresponding pattern in which the status of Cues A–C was switched with the status of Cues D–F. For instance, the probability of a patient having Flu Strain 1 given Cue Pattern abc DeF was matched with the probability of the patient having Flu Strain 1 given AbC def (cue patterns that would be matched with themselves [e.g., AbC DeF] were not included in the analyses to avoid inflating the apparent correspondence between the sets of cues). We would expect a high correlation between the probability pairs to the extent that Cues A–C and D–F have similar diagnostic implications. Using this method, the correlation between the probabilities is $r = .9998$, and the square root of the mean squared difference between each pair of probabilities is 0.8% on a 0%–100% scale. In short, the diagnostic value of the two sets of cues appears to be well equated no matter which metric is used.

4. The normality of the distribution of R_{unk} when the critical cues were high present-frequency cues could be questioned because the mean and median values are different (.82 and .52, respectively). The difference between the mean and median values is mainly caused by one extreme value of $R_{unk} = 9.75$ for 1 participant when the critical cues were high present-frequency cues. When this participant's data is removed, the mean (.64) and median (.51) are fairly similar while the difference between R_{unk} for the high and low present-frequency cues remains unchanged [respectively: means = .64 and .18; SEMs = .15 and .14; $F(1,51) = 4.17$, $MS_e = 1.31$, $p < .05$]. As in the previous experiments, the distribution of R_{unk} approximates a normal distribution quite well when the data from this participant is removed.