Title
Low Dimensional Representations in Multi-Cue Judgment

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Zhao, Joyce Wenjia
Bhatia, Sudeep
Davis-Stober, Clintin P.

Publication Date
2017

Peer reviewed

Low Dimensional Representations in Multi-Cue Judgment

Joyce Wenjia Zhao (zhaowenj@sas.upenn.edu)
Department of Psychology, University of Pennsylvania
Philadelphia, PA Address

Sudeep Bhatia (bhatiasu@sas.upenn.edu)
Department of Psychology, University of Pennsylvania
Philadelphia, PA Address

Clintin P. Davis-Stober (stoberc@missouri.edu)
Department of Psychological Sciences, University of Missouri
Columbia, MO

Abstract

The study of multi-cue judgment investigates how decision makers integrate cues to predict the value of a criterion variable. We consider a multi-cue judgment task in which decision makers have prior knowledge of inter-cue relationships but are ignorant of how the cues correlate with the criterion. In this setting, a naive judgment strategy prescribes an equal weight for each cue. However, we find that many participants appear to use a weighting scheme based on a low-dimensional representation of the cue space. The use of such a representation is consistent with core insights in semantic memory research and has important optimality properties concerning judgment accuracy.

Keywords: judgment and decision making; cue integration; improper linear models; dimensionality reduction; semantic memory

Introduction

Effective judgment and decision making involves the aggregation of multiple cues, or pieces of information, to evaluate a criterion variable. For example, individuals may receive advice from two or more friends regarding a financial investment, and aggregate this advice to calculate the expected return on the investment. Alternatively, they may have to choose between job candidates with multiple attributes, and have to aggregate these attributes to determine the quality of the candidates.

Traditionally, many normative and descriptive models of judgment and decision making adopt a linear approach and propose that decision makers compute the value of the criterion using a weighted average of the cues, with the weights being proportional to the observed relationship between the cues and the criterion (Brunswik, 1952; Keeney & Raiffa, 1993). Linear models are often criticized as they require large amounts of information and abundant cognitive resources in order to be accurate. Thus, many researchers have proposed that individuals use *improper* linear models, such as heuristics. These models involve a fixed weighting scheme that assigns a priori weights to the cues. For example, an equal weights model gives each cue the same weight, and the lexicographic model assigns all the weight to a single cue (Dawes, 1979; Gigerenzer & Todd 1996). These models have been shown to perform as well as, if not better than, proper linear models in many situations, ranging from graduate student admission to clinical predictions (Dawes, 1979).

In addition to being cognitively simpler, improper linear models can also be used in situations where proper linear models are inapplicable. Consider settings where the relationship between the cues and the criterion is completely unknown. For example, individuals using the advice of their friends to judge an investment may not have previously observed how well their friends predict the performance of such investments. Likewise, individuals evaluating job candidates for novel or unconventional jobs may have never observed the value of different candidate attributes in the context of these jobs. In these situations, decision makers may have detailed knowledge about the relationship between the cues (e.g. how often their friends agree with each other or how frequently job candidate attributes co-occur) but have no way to assign weights to the cues in accordance with the standard linear model (where weights depend on the cues' relationship with the criterion). However, an a priori weighting scheme, as proposed by improper linear models, can still be used to make an evaluation.

For multi-cue judgment with known inter-cue relationships, but unknown cue-criterion relationships, the key questions of interest are the following: Which improper weighting scheme *should* decision makers use and which schemes *do* decision makers use. The former question has been tackled by Davis-Stober, Dana & Budescu (2010a, 2010b). Davis-Sober et al. propose that any possible weighting scheme, $\boldsymbol{\beta}$, can be assessed with regards to how far it deviates from the true weight vector, $\boldsymbol{\beta}^*$, by taking the sum of squared difference between the weights in $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$, i.e. $\sum_i(\beta_i - {\beta_i}^*)^2$. When the cue-criterion relationships are unknown, $\boldsymbol{\beta}^*$ is also unknown. In these settings optimizing $\boldsymbol{\beta}$ can be seen as involving minimizing the *risk*, defined as the expectation of sum of squared error of $\boldsymbol{\beta}$, $\sum_i(\beta_i - {\beta_i}^*)^2$. By this standard, the best improper linear weighting scheme is the eigenvector corresponding to the first (i.e. largest) eigenvalue of the inter-cue correlation matrix (see Davis-Stober et al. 2010 for details). We will refer to this weighting scheme as $\boldsymbol{\beta}_{EV1}$. Here *EV1* in the subscript refers to the use of the eigenvector corresponding to the first eigenvalue.

$\boldsymbol{\beta}_{EV1}$ depends on the relationship between the cues in the judgment task, and can be shown, in appropriate settings, to approximate other existing improper linear models. For example, if the cues are equally, and positively, correlated with each other, $\boldsymbol{\beta}_{EV1}$ assigns an equal weight to each cue, similar to the equal weights model. In contrast, if cue 1 is highly correlated with all the other cues, and all the other cues are moderately correlated or uncorrelated internally, $\boldsymbol{\beta}_{EV1}$ overweighs cue 1 relative to other cues. This can mimic a lexicographic judgment strategy.

Normative solutions aside, descriptively, which weighting scheme do decision makers actually use when integrating multiple cues with unknown cue-criterion relationships? A first guess involves an equal weights model: Without knowing which cues are more related to the criterion than others, it seems conceivable that decision makers assign the same weights to all the available cues. This corresponds to a type of ignorance prior. However, a more principled guess could rely on insights regarding semantic representation. Decision makers with prior experience with the cues may have learnt mental representations of the cues. These representations, in many settings, correspond to projections of the decision makers' experiences with the cues onto a low-dimensional space. Such projections can be approximated by a principle components analysis on the cue-correlation matrix, or equivalently, a singular value decomposition on the matrix of cue-context co-occurrence. Indeed, such a decomposition is a key component of numerous existing approaches to modelling semantic representation, including latent semantic analysis (Landauer & Dumais, 1997), multi-dimensional scaling (Kruskal & Wish, 1978), and neural network models of semantic memory (Saxe, McClelland & Ganguli, 2013). Interestingly, such a decomposition also yields the normative $\boldsymbol{\beta}_{EV1}$ model when only the first latent dimension of the projection is used to evaluate the criterion.

The goal of this paper is to investigate the plausibility of the $\boldsymbol{\beta}_{EV1}$ weighting scheme, and to compare its ability to predict participant judgments with alternate improper linear models such as the equal weights rule and the lexicographic rule. To appropriately test these models, we examine settings where participants have prior knowledge of inter-cue correlations but do not know how the different cues correlate with the criterion in consideration. Additionally, we systematically vary the cue-correlation matrix, and subsequently $\boldsymbol{\beta}_{EV1}$, in order to adequately differentiate the predictions of this weighting scheme from those of alternate weighting schemes in our studies. We demonstrate the applicability of $\boldsymbol{\beta}_{EV1}$ for describing participant behavior in two ways: 1) by examining the model fits for $\boldsymbol{\beta}_{EV1}$ relative to other improper linear models, and 2) by testing whether the weights assigned by $\boldsymbol{\beta}_{EV1}$ predict decision makers' use of these other improper models.

General Method

In our three studies, the multi-cue judgment task was presented as an advice integration task, with the cues in consideration corresponding to the judgments of four advisors (similar to Bröder, 2003). They were described as predicted stock prices in Studies 1 and 2 and restaurant ratings in Study 3. Correspondingly, the criterion was the true stock price in Studies 1 and 2 and the true restaurant quality in Study 3. The cue-criterion correlations were never revealed to the participants.

The studies consisted of three tasks. The first two tasks exposed the participants to the cues, so as to allow them to form mental representations of the cue space. The third task asked participants to predict the criterion value based on the cues. In addition to being stated numerically, cue values in the three tasks were also shaded based on their magnitude. Participants were told that the cue values ranged from 0-100 and were all centered at 50. They were also told that some cues (advisors) might be more similar to each other, and that it was useful to pay attention to how closely different cues agreed with each other.

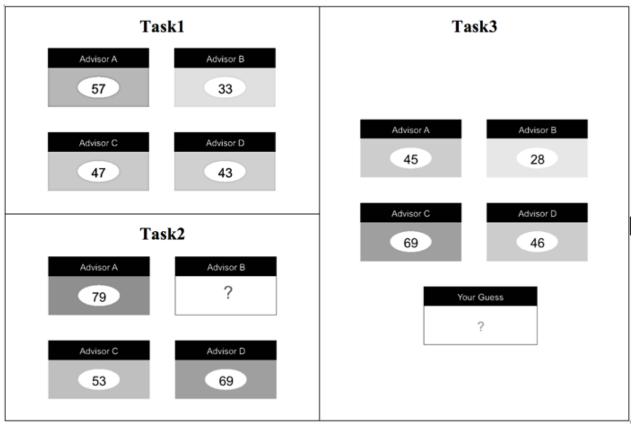

Figure 1. Stimuli display for Task 1-3.

In task 1, participants saw the four cues in 25 trials (Figure 1: upper left) displayed in four boxes. Each trial presented a set of cue values, and participants were asked to merely observe the cue values, without providing a response. In task 2, participants continued to learn the cue values, this time with feedback. Particularly, only three of the four cues were shown to participants (Figure 1: lower left). Participants had to guess the value of the fourth cue based on their knowledge of the inter-cue relationships. After the participant's guess, the real cue value was revealed. To increase motivation, participants were provided with a summary of their performance accuracy after every 50 trials. The cue to be guessed was determined at random in each trial.

In task 3, participants were shown all four cue values, and were asked to make a guess regarding the value of the criterion (real stock price for Studies 1 and 2 and actual restaurant quality for Study 3; Figure 1: right). The true value of the criterion was not revealed after participants' guesses, so that participants stayed uninformed regarding the cue-criterion relationship. Task 3 was the most relevant to our research question, as it provided a direct test of how cue values were integrated to make a judgment of the criterion.

Study 1

Study 1 examines the predictions of the $\boldsymbol{\beta}_{EV1}$ weighting rule by considering a setting in which inter-cue relationships lead to a larger weight on one cue and smaller weights on the remaining cues.

Methods

44 participants (37 females; Mean Age = 19.7, SD Age = 1.2), recruited from a university experimental participation pool, completed this study in a behavioral laboratory.

The study involved a hypothetical stock prediction task. The cue values were stock prices predicted by four advisors. For each cue, the values were normally distributed, with a mean of 50 and a standard deviation of 25. The inter-cue correlation matrix of the advisors is shown in Figure 2a.

(a)

	$C1$	$C2$	$C3$	$C4$
$C1$	1			
$C2$	.6	1		
$C3$	.6	.05	1	
$C4$	.6	.05	.05	1

(b)

	$C1$	$C2$	$C3$	$C4$
$C1$	1			
$C2$	.4	1		
$C3$	.4	.4	1	
$C4$	.4	.4	.4	1

(c)

	$C1$	$C2$	$C3$	$C4$
$C1$	1			
$C2$	.8	1		
$C3$	.1	.1	1	
$C4$	.1	.1	.4	1

Figure 2 (a) Inter-cue correlation matrix for Study 1 and the treatment condition of Study 2. (b) Inter-cue correlation matrix for the control condition of Study 2. (c) Inter-cue correlation matrix for Study 3.

As can be seen in this matrix, cue 1 is highly correlated with all the three other cues, with a correlation coefficient of 0.6. The internal correlation among the remaining cues is very weak, with a correlation coefficient of 0.05. The eigenvector corresponding to the first (largest) eigenvalue of the cue correlation matrix is $\boldsymbol{\beta}_{EV1} = [0.35, 0.22, 0.22, 0.22]$. Using this weighting vector, leads to an overweighting of cue 1, and a relative underweighting of the remaining cues.

We used the above distributions to generate a single set of stimuli for all participants, for tasks 1, 2 and 3. For each participant, the display position for each of the four cues was randomly chosen at the beginning of the study and stayed unchanged for the entire session. In other words, the specific advisor (advisor A, B, C or D) associated with cue 1, was counterbalanced.

Results

We first examined participants' performance in task 2, where they used three cue values for guessing the remaining cue value. Our analysis of behavior in this task suggested that participants were able to successfully learn the underlying cue structure. Particularly they placed a higher weight on cue 1 relative to the other cues when predicting the remaining cues ($p < 0.001$). Due to space constraints we will not outline these results in more detail (they will be reported elsewhere).

We also investigated the weighting scheme used by participants when integrating cues to predict criterion values in task 3. For this purpose, we considered a number of candidate weighting schemes, including $\boldsymbol{\beta}_{EV1}$ (corresponding to the first eigenvector of the cue-correlation matrix) and $\boldsymbol{\beta}_{EW}$ (corresponding to the equal weighting rule). We also considered lexicographic rules. Here, we tested four models that put all the weights on a single cue. These were referred to as $\boldsymbol{\beta}_{LEX1}$, $\boldsymbol{\beta}_{LEX2}$, $\boldsymbol{\beta}_{LEX3}$ and $\boldsymbol{\beta}_{LEX4}$, corresponding to the cue that was given the unit weight. In addition to $\boldsymbol{\beta}_{EV1}$, we also considered the linear weighting schemes corresponding to the remaining three eigenvectors of the cue correlation matrix. These are referred to as $\boldsymbol{\beta}_{EV2}$, $\boldsymbol{\beta}_{EV3}$ and $\boldsymbol{\beta}_{EV4}$. Therefore, we have in total nine improper linear weighting schemes to compare. Each linear weighting scheme defined a weighting vector for the four cues. E.g. $\boldsymbol{\beta}_{EV1} = [0.35, 0.22, 0.22, 0.22]$, $\boldsymbol{\beta}_{EW} = [0.25, 025, 0.25, 0.25]$, $\boldsymbol{\beta}_{LEX1} = [1, 0, 0, 0]$, etc. For comparability, the weights in each scheme were constrained to add up to one.

In order to scale the criterion estimate generated by these improper linear models to match the participants' guesses, we introduced two additional participant-level parameters, α_0 and α_1, so that the predicted guess for each weighting scheme was $\alpha_0 + \alpha_1 \boldsymbol{\beta} \cdot \boldsymbol{C}$. Here $\boldsymbol{\beta}$ corresponds to the weighting vector of the model in consideration, and $\boldsymbol{C}$ is the vector of cue values presented in the trial. We also assumed a normally distribute error, with standard deviation σ, and subsequently fit each of these nine models by maximizing log-likelihood. α_0, α_1 and σ were allowed to vary across the nine models. The model fitting was done on the participant level. Because the linear weighting schemes were pre-determined, each model used same number of parameters (3 parameters: α_0, α_1 and σ) to predict each participant's 100 guesses in task 3.

We compared participant level log likelihood values for the nine candidate models (Table 1). Since all models have the same number of parameters, our model comparison is equivalent to model selection by AIC. Among 44 participants, 10 participants' predictions were best described by $\boldsymbol{\beta}_{EV1}$, 23 by $\boldsymbol{\beta}_{EW}$, 8 by $\boldsymbol{\beta}_{LEX1}$, 1 by $\boldsymbol{\beta}_{EV2}$, 1by $\boldsymbol{\beta}_{EV3}$, and 1 by $\boldsymbol{\beta}_{LEX4}$. When comparing only $\boldsymbol{\beta}_{EV1}$ and $\boldsymbol{\beta}_{EW}$, 19 participants were better described by $\boldsymbol{\beta}_{EV1}$, whereas 25 were better described by $\boldsymbol{\beta}_{EW}$. According to a paired Wilcoxon test on participant level model fits, there was no significant difference between log likelihood values of the $\boldsymbol{\beta}_{EV1}$ model ($Median = -351.38$) and the $\boldsymbol{\beta}_{EW}$ model ($Median = -351.59$), $Z = 1.24, p = 0.216$. Paired Wilcoxon tests also indicated that fits for all of the remaining models were significantly worse than those for the $\boldsymbol{\beta}_{\boldsymbol{EV1}}$ model and $\boldsymbol{\beta}_{\boldsymbol{EW}}$ model (p < 0.001).

Table 1: Comparison of model fits for Study 1

	Parameter (Median)			Log Likelihood		
Model	α_0	α_1	σ	Median	Mean	# best
EV1	7.32	0.84	8.57	-351.38	-357.16	10
EV2	49.92	0.02	18.66	-424.01	-420.56	1
EV3	49.96	0	18.63	-423.86	-420.64	1
EV4	45.37	0.32	18.63	-424.05	-420.6	0
EW	4.94	0.89	8.56	-351.59	-357.10	23
LEX1	17.94	0.63	8.76	-351.39	-358.54	8
LEX2	29.81	0.37	15.83	-408.81	-406.45	0
LEX3	32.17	0.38	15.95	-410.32	-407.24	0
LEX4	29.7	0.39	15.58	-408.7	-405.83	1

Although participants had no information regarding the validities of any cues, a substantial subgroup of participants did not simply assign equal weights to cues. Instead, they overweighed cue 1 as predicted by $\boldsymbol{\beta}_{\boldsymbol{EV1}}$. The fact that some participants were actually best fit by $\boldsymbol{\beta}_{\boldsymbol{LEX1}}$ suggested that some participants overweighed cue 1 even more than $\boldsymbol{\beta}_{\boldsymbol{EV1}}$ recommended. Only one participant was best described by the other three lexicographic rules, indicating that $\boldsymbol{\beta}_{\boldsymbol{EV1}}$ can predict which single cue participants tend to overweigh.

Study 2

In Study 1, we found that the EV1 and EW models described participant level data about equally well, in terms of average log likelihood and proportion best fit. However, it is possible that predictions made by EV1 and EW were similar enough to be practically indistinguishable (given the noise in the data). This could confound our interpretation of model fit. Study 2 addresses this alternative explanation by manipulating the inter-cue relationships between subjects.

Methods

64 participants (35 females; Mean Age = 19.9, SD Age = 1.1), recruited from a university experimental participation pool, completed this study in a behavioral laboratory.

All aspects of the study design were kept identical to Study 1, except that the cue correlation matrix varied between a treatment condition and a control condition. Participants were randomly assigned to one of these two conditions at the start of the study.

For the treatment condition, the inter-cue correlation matrix was identical to that in Study 1 (Figure 2a), generating an optimal weighting scheme with $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat} =$ [0.35, 0.22, 0.22, 0.22] (here we use the superscript to distinguish the treatment vs. control condition). For the control condition, the cue correlation matrix kept the correlation between all the cues constant at 0.4 (Figure 2b). Therefore, the weighting vectors predicted by the optimal weighting scheme and the equal weights rule were both $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Cont} = \boldsymbol{\beta}_{EW} =$ [0.25, 0.25, 0.25, 0.25]. Due to different inter-cue relationships across conditions, $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ should provide a better account of behavior in the treatment condition compared to the control condition. Likewise $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Cont} = \boldsymbol{\beta}_{EW}$ should provide a better account of behavior in the control condition compared to the treatment condition (even if a large subgroup of participants in the treatment condition do place an equal weight on all cues).

Results

31 participants were assigned to the treatment condition and 33 participants were assigned to the control condition. As in Study 1, we first looked at participant learning in task 2. In the treatment condition, participants did learn the special status of cue 1. Particularly, as in Study 1, they placed a higher weight on cue 1 relative to the other cues when predicting the remaining cues ($p < 0.001$). In the control condition, participants placed similar weights on cues 1-4 when predicting cue values, indicating that they learnt different inter-cue relationships for the two conditions. Manipulating the cue-correlation matrix thus had an effect on participant learning. This laid the basis for task 3, where participants integrated cues to predict criterion values (again, due to space constraints, we will not expand on these results here).

Next, we examined which weighting schemes were used by participants in task 3. For both conditions, we applied the model fitting procedures of Study 1, and nine linear weighting schemes were compared on the participant level (Tables 2 and 3). Out of 31 participants in the treatment condition, 6 were best described by $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$, 18 by $\boldsymbol{\beta}_{\boldsymbol{EW}}$, 6 by $\boldsymbol{\beta}_{\boldsymbol{LEX1}}$ and 1 by $\boldsymbol{\beta}_{\boldsymbol{EV2}}$. $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ outperformed $\boldsymbol{\beta}_{\boldsymbol{EW}}$ for a substantial subgroup of participants (13 out of 31). As in Study 1, some participants were best described by $\boldsymbol{\beta}_{\boldsymbol{LEX1}}$, indicating that they overweighed cue 1 more than recommended by $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$. No participant was best described by the other three lexicographic rules, indicating that $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ can predict decision makers' use of other improper linear models in the treatment condition.

We also compared the log likelihood values of the fits. Although the log likelihood values of the $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ model were significantly smaller than those of the $\boldsymbol{\beta}_{\boldsymbol{EW}}$ model ($Z = -2.06, p = 0.040$), the effect size was small ($Median_{EV1} = -350.56$, $Median_{EW} = -350.17$). Additionally both the $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ model and $\boldsymbol{\beta}_{\boldsymbol{EW}}$ model predicted participant level data significantly better than all other models ($p < 0.001$). These results replicate findings of Study 1.

In the control condition, the inter-cue correlation matrix was balanced and the weighting schemes for $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Cont}$ and $\boldsymbol{\beta}_{\boldsymbol{EW}}$ were identical. Unsurprisingly, all 33 participants were better described by $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Cont} = \boldsymbol{\beta}_{EW}$ than any other models (Table 3). The fact that no participants were best fit by lexicographic rules in the control condition but some were best fit by $\boldsymbol{\beta}_{\boldsymbol{LEX1}}$ in the treatment condition again indicated that participants' cue weighting behavior can be predicted by the inter-cue correlation matrix.

Lastly, we examined the predictions of $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ on the data from the control condition. For this purpose we fit a tenth model in the control condition, with weights given by $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ (and α_0 , α_1 and σ flexible). Unlike the treatment condition, this model outperformed the $\boldsymbol{\beta}_{EW}$ = $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Cont}$ model for only 4 out of 33 participants in the control condition. A paired Wilcoxon test indicated that the log likelihoods of the $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ model on the control-condition data ($Median = -336.73$) were significantly lower than those of the $\boldsymbol{\beta}_{EW}$ = $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Cont}$ model ($Median = -335.76$), $Z = -4.24, p < 0.001$.

Table 2: Comparison of model fits for Study 2 (treatment)

	Parameter (Median)			Log Likelihood		
Model	α_0	α_1	σ	Median	Mean	#best
EV1	7.70	0.90	10.07	-350.56	-363.55	6
EV2	50.65	0.03	19.73	-424.88	-423.34	1
EV3	50.72	0.01	19.73	-425.08	-423.60	0
EV4	46.79	0.33	19.66	-424.85	-423.52	0
EW	5.10	0.95	10.07	-350.17	-363.35	18
LEX1	19.28	0.67	10.19	-353.77	-365.33	6
LEX2	31.88	0.40	16.28	-407.65	-408.75	0
LEX3	33.68	0.39	17.28	-412.35	-411.83	0
LEX4	30.22	0.44	16.59	-411.66	-409.48	0

Table 3 Comparison of model fits for Study 2 (control)

	Parameter (Median)			Log Likelihood		
Model	α_0	α_1	σ	Median	Mean	#best
EV1/EW	10.59	0.83	7.49	-335.76	-336.13	29
EV2	51.41	0.00	17.64	-415.22	-411.91	0
EV3	51.41	0.00	17.64	-415.22	-411.92	0
EV4	51.39	0.07	17.63	-414.84	-411.69	0
LEX1	28.91	0.45	13.00	-387.19	-385.25	0
LEX2	30.35	0.43	13.30	-391.31	-388.82	0
LEX3	28.55	0.46	12.20	-382.23	-382.13	0
LEX4	31.17	0.44	13.32	-390.59	-387.50	0
EV1Treat	11.04	0.81	7.75	-336.73	-340.25	4

Overall, the differences in the mean and median log likelihoods of the $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Treat}$ and the $\boldsymbol{\beta}_{\boldsymbol{EV1}}^{Control}$ = $\boldsymbol{\beta}_{EW}$ models in the control condition were 4.12 and 0.97 respectively. These were larger than the equivalent differences in the treatment condition, which were 0.20 and 0.39 (these differences were 0.06 and -0.21 in Study 1). These results indicate that the relatively good fits for the $\boldsymbol{\beta}_{\boldsymbol{EV1}}$model in the treatment condition of Study 2 and in Study 1 were not due to this model mimicking the equal weights rule.

Study 3

Study 3 provides a more stringent test of the $\boldsymbol{\beta}_{\boldsymbol{EV1}}$model by considering a setting with more complex inter-cue relationships. It also examines judgments of restaurant quality rather than stock performance.

Methods

46 participants (34 females; Age Mean = 19.3, SD Age = 1.0) recruited from a university experimental participation pool, completed this study in a behavioral laboratory.

The study was framed as involving judgments of restaurant quality. Here the cue values were restaurant scores rated by four reviewers, and the criterion corresponded to the real restaurant quality. Other aspects of the study design were kept identical to Study 1, except the inter-cue correlation matrix, which was changed to the matrix displayed in Figure 2c. Here cue 1 is highly correlated with cue 2, cue 3 is moderately correlated with cue 4, and cues 1 and 2 are weakly correlated with cues 3 and 4. With this inter-cue correlation structure, $\boldsymbol{\beta}_{\boldsymbol{EV1}}$ predicts a weighting vector of [0.35, 0.35, 0.15, 0.15], i.e. an overweighting of cues 1 and 2, relative to 3 and 4.

Results

As in Studies 1 and 2, our analysis of behavior in task 2 suggested that participants were able to successfully learn the underlying cue structure. Particularly they relied more on cues 1 and 2 than on cue 3 and 4 when guessing for cues 1 and 2; they also relied more on cues 3 and 4 than on cues 1 and 2 when guessing for cues 3 and 4. Again, due to space constraints we will not outline these results in more detail.

Next, we investigated the linear weighting schemes used by participants in task 3. The nine candidate weighting schemes and the model fitting procedures were the same as in Study 1 (though, of course, $\boldsymbol{\beta}_{\boldsymbol{EV1}}$, $\boldsymbol{\beta}_{\boldsymbol{EV2}}$, $\boldsymbol{\beta}_{\boldsymbol{EV3}}$, $\boldsymbol{\beta}_{\boldsymbol{EV4}}$assigned different weights to the cues in Study 3, compared to the corresponding models in Study 1).

Out of the 46 participants, 7 were best fit by $\boldsymbol{\beta}_{\boldsymbol{EV1}}$and 38 were best fit by $\boldsymbol{\beta}_{\boldsymbol{EW}}$. The remaining participant was best fit by $\boldsymbol{\beta}_{\boldsymbol{LEX2}}$. When comparing only $\boldsymbol{\beta}_{\boldsymbol{EV1}}$and $\boldsymbol{\beta}_{\boldsymbol{EW}}$, we found that 8 participants were better described by $\boldsymbol{\beta}_{\boldsymbol{EV1}}$than $\boldsymbol{\beta}_{\boldsymbol{EW}}$. Additionally, $\boldsymbol{\beta}_{\boldsymbol{EW}}$ ($Median = -350.82$) was significantly better than $\boldsymbol{\beta}_{\boldsymbol{EV1}}$ ($Median = -353.15$) according to a paired Wilcoxon test performed on participant level log likelihood values, $Z = 3.98, p < 0.001$. Except for $\boldsymbol{\beta}_{\boldsymbol{EW}}$, $\boldsymbol{\beta}_{\boldsymbol{EV1}}$outperformed all the other candidate models (with $p < 0.001$). As would be predicted by the $\boldsymbol{\beta}_{\boldsymbol{EV1}}$model, the lexicographic models did not provide a good account of participant behavior in this study. Table 4 provides additional details regarding these fits.

Table 4 Comparison of model fits for Study 3

	Parameter (Median)			Log Likelihood		
Model	α_0	α_1	σ	Median	Mean	# best
EV1	14.04	0.76	9.28	-353.15	-358.26	7
EV2	44.77	0.26	16.76	-412.07	-410.59	0
EV3	51.04	0.10	17.29	-415.25	-413.63	0
EV4	50.95	0.03	17.35	-415.29	-413.82	0
EW	6.32	0.88	8.75	-350.82	-349.09	38
LEX1	30.22	0.44	12.58	-383.14	-384.67	0
LEX2	28.31	0.49	11.94	-379.23	-381.29	1
LEX3	32.89	0.34	15.32	-403.08	-401.97	0
LEX4	32.01	0.37	14.63	-400.57	-398.18	0

As in previous studies, we found that a significant subgroup of participants overweighed some cues (as suggested by $\boldsymbol{\beta_{EV1}}$), rather than simply averaging all the available cues (as suggested by $\boldsymbol{\beta_{EW}}$). That said, the performance of $\boldsymbol{\beta_{EV1}}$ was relatively worse in this study compared to our previous studies. This could be due to the differences in the cue-correlation matrices, suggesting that decision makers are less likely to use the $\boldsymbol{\beta_{EV1}}$ scheme when the underlying cue structure is complex. These differences could also, however, be attributed to the change in the task frame. Restaurant quality is more subjective than stock performance, and decision makers may be less likely to rely on the cue-correlation structure in these subjective settings.

Discussion

In three studies, we investigated how decision makers weigh cues when cue criterion relationships are unknown. The optimal improper linear model uses the eigenvector, $\boldsymbol{\beta_{EV1}}$, corresponding to the largest eigenvalue of the cue correlation matrix (Davis-Stober et al., 2010a, 2010b). Low dimensional representations of the cue space, learnt by some common models of semantic memory (Kruskal & Wish, 1978; Landauer & Dumais, 1997; Saxe et al., 2013), can also produce this type of weighting scheme.

Our results suggest that $\boldsymbol{\beta_{EV1}}$ provides a good description of participants' behavior. This model outperformed all other improper linear models tested in this paper, except for the equal weights model (with weights $\boldsymbol{\beta_{EW}}$). On the aggregate level, the log likelihoods for the $\boldsymbol{\beta_{EV1}}$ and $\boldsymbol{\beta_{EW}}$ weighting scheme were relatively close, showing no meaningful differences in Study 1, very minor differences in the treatment condition of Study 2, and somewhat larger differences in Study 3. As for individual level fits, there existed a substantial group of participants for whom $\boldsymbol{\beta_{EV1}}$ outperformed $\boldsymbol{\beta_{EW}}$. The size of this group ranged from 43% of the participant pool in Study 1, 42% in the treatment condition of Study 2, and 17% in Study 3. Moreover, a comparison of the control and the treatment conditions of Study 2 showed that experimental manipulations that varied the inter-cue correlation matrix influenced relative model fits.

$\boldsymbol{\beta_{EV1}}$ was also able to predict when and how participants used lexicographic weights. When $\boldsymbol{\beta_{EV1}}$ prescribed equal weights (control condition of Study 2) or the overweighing two cues (Study 3), there were almost no participants who were best described by such lexicographic weighting schemes. In contrast, in Study 1 and the treatment condition of Study 2, $\boldsymbol{\beta_{EV1}}$ overweighed a single cue. In these conditions, a substantial group of participants (18% in Study 1 and 19% in the treatment condition of Study 2) behaved according to a lexicographic rule that placed all of the weight on this cue (in contrast lexicographic rules that prioritize other cues all performed very poorly).

That said, $\boldsymbol{\beta_{EV1}}$ did not provide a good account of behavior in Study 3, which adopted a more complex inter-cue correlation matrix. The results of this study suggest that such a weighting scheme may not be used in all settings. Additionally, the equal weights rule was the majority model in all studies, indicating that most participants tend to use the simpler equal weights strategy (corresponding to an ignorance prior) in the absence of cue-criterion knowledge. Further work should examine the effect of inter-cue correlation structure and individual differences on the use of the $\boldsymbol{\beta_{EV1}}$ weighting rule. This work may extend the insights of other cognitive models of multi-cue judgment, such as those relying on neural network representations (Glöckner, Hilbig & Jekel, 2014) or exemplar memory-based predictions (Juslin, Karlsson & Olsson, 2008). Such models have not been applied to settings in which cue-criterion relationships are unknown. However, they nonetheless provide formal predictions regarding the learning and representation of cue knowledge and its relationship with the statistical structure of the judgment environment. For this reason they may provide a more adequate framework for understanding the cognitive underpinnings of the $\boldsymbol{\beta_{EV1}}$ weighting model.

References

Bröder, A. (2003). Decision making with the" adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4), 611.

Brunswik, E. (1952). The conceptual framework of psychology. Psychological Bulletin, 49(6), 654-656.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571.

Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010). Why recognition is rational: Optimality results on single-variable decision rules. *Judgment and Decision Making*, 5(4), 216.

Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010). A constrained linear estimator for multiple regression. *Psychometrika*, 75(3), 521-541.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In Simple heuristics that make us smart (pp. 3-34). Oxford University Press.

Glöckner, A., Hilbig, B. E., & Jekel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition*, 133(3), 641-666.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, 106(1), 259-298.

Keeney, R. L., & Raiffa, H. (1993). Decisions with multiple objectives: preferences and value trade-offs. Cambridge university press.

Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling (Vol. 11). Sage.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Learning hierarchical category structure in deep neural networks. In *Proceedings of the 35th annual meeting of the Cognitive Science Society* (pp. 1271-1276).