This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Technical Correspondence

A LAMSTAR Network-Based Human Judgment Analysis

Jae M. Yoon, David He, and Matthew L. Bolton, *Senior Member, IEEE*

***Abstract*—Judgment analysis (JA) is a technique for modeling and interpreting human judgments that is usually based on multiple linear regression. However, the linear assumptions inherent to this approach can be limiting for modeling both the human judgments and the environmental criterion. This paper addresses this by introducing a formulation of JA based on large memory storage and retrieval (LAMSTAR) artificial neural networks. We describe our LAMSTAR network JA process and use it to analyze data from an air traffic control conflict prediction task. These results are compared with those of a traditional regression-based lens model analysis. We found that the LAMSTAR-based JA did a better job of capturing human judgment, while the regression-based model was more appropriate for the criterion. This suggest that the LAMSTAR-based JA approach has utility when human judgments are not well represented by a linear model. We discuss our results with respect to both the specific application we evaluated as well as meta-analyses of the JA literature. We also explore avenues for future research.**

***Index Terms*—Judgment, lens model, linear regression, neural networks.**

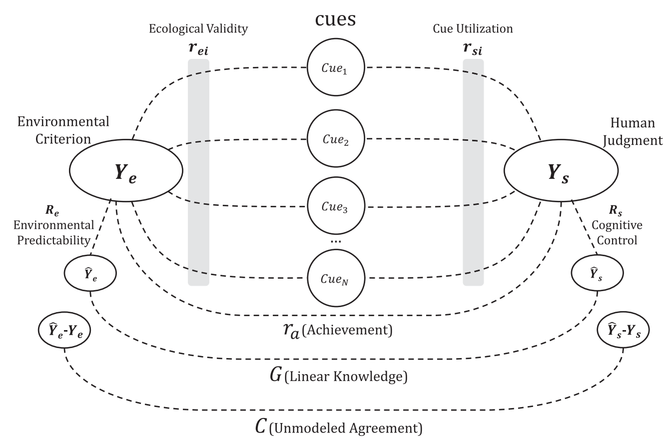

Fig. 1. Graphical representation of the double-system lens model. Note that i is an integer in the range [1, N].

I. Introduction

Human judgment represents a person's assessment of an attribute or quality from the environment based on the available ecological information. Because judgment occurs in many human activities, its accuracy can be critical to a human's ability to successfully achieve goals when interacting with their environment. Thus, being able to understand and predict human judgment can provide engineers and analysts with information that can help them analyze, design, and evaluate systems that rely on humans for their safe and efficient operation. As such, attempts have been made to mathematically model human judgment as a transformation of information from the environment. The most successful of these is judgment analysis (JA). JA typically uses multiple linear regression to model and interpret human judgments and compare them to an environmental criterion. However, the linear assumptions inherent to this approach can be limiting. This paper explores the development of a JA technique based on large memory storage and retrieval (LAMSTAR) artificial neural networks (ANNs) that do not require the linear assumptions of typical judgment analyses. In the remainder of this section, we discuss JA, ANNs, and LAMSTAR networks to motivate the development of our new JA approach.

A. Judgment Analysis

JA, which is based on Brunswiks probabilistic functionalism [1], [2], describes human judgments as an ecological response to the environment, mediated by cues. To date, applications of JA have mostly used multiple linear regression (via multiple specific formulations [2]) to create a number of measures of judgment performance. In this work, we are primarily concerned with the double-system lens model [2] (see Fig. 1).

The double-system lens model, which is employed frequently in the literature [3], [4], uses symmetric statistical models between the environment and judgment to describe the judgment process. Specifically, the same measurable environmental cues are used as predictors (independent variable) in two fitted linear regression models: one to the criterion (the actual value of the environmental quality that is being judged; $\hat{Y}_e$) and one to judgment values ($\hat{Y}_s$). The weights assigned to the cues (independent variables) allow analysts to compare how differently the cues factor into the prediction of the criterion (ecological validities) and the judgment (cue utilizations). The cue utilizations can be compared between participants to determine how the judgment strategies of each differ. In cases where the cue data have been normalized, cue weights can also be used to compare the relative weight each cue has in influencing a predicted dependent measure (criterion or judgment). Further, the lens model equation [5], [6]

$$r_a = G \cdot R_e \cdot R_s + C \cdot \sqrt{1 - R_e^2} \cdot \sqrt{1 - R_s^2} \qquad (1)$$

gives analysts a means of evaluating the achievement of the judge while accounting for the different factors that affect it. r_a is the achievement of the judge represented as the correlation between the criterion (Y_e) and judgment (Y_s). Thus, achievement is measured from low to high by a value between 0 and 1. G represents linear knowledge, a measure of how the model predictions between the environment and the human correspond. This is measured as the correlation between $\hat{Y}_e$ and $\hat{Y}_s$. R_e is a measure of the environmental predictability, how well the model of the environment corresponds to the environmental criterion, measured as a correlation between Y_e and $\hat{Y}_e$. Similarly, R_s represents cognitive control in that it is a measure of how well the human judgment model

Manuscript received January 7, 2016; revised June 29, 2016 and August 11, 2016; accepted September 17, 2016. This paper was recommended by Associate Editor B. Donmez.

J. M. Yoon and D. He are with the University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: alexxyoon@gmail.com; davidhe@uic.edu).

M. L. Bolton is with the Department of Industrial and Systems Engineering, University at Buffalo, Amherst, NY 14260-2050 USA (e-mail: mbolton@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2016.2612231

matches the actual human judgment (the correlation between Y_s and $\hat{Y}_e$). Finally, C represents unmodeled agreement: a measure of the correspondence between the information not captured between the two models. This is measured as the correlation between the residuals of the environment model ($\hat{Y}_e - Y_e$) and the judgment model ($\hat{Y}_s - Y_s$).

In this form, JA has been successfully used to model and evaluate human judgment in a number of domains including policy making [7], [8], medicine [9], weather forecasting [10], education [11], and air traffic control [12]–[14]. Despite this success, there are several limitations to this approach. This technique assumes that human judgments and the criterion can be reasonably molded as a linear combination of weighted ecological cues and, thus, satisfy all of the requirements for multiple linear regression. While this can result in good models for both the human and the environment in many situations [3], [4], [15], [16], there can be exceptions for both. First, there are cases where human judgment is based on a nonlinear synthesis of environmental information. This can include situation where people are under time pressure, they are using intuition, they must consider multiple alternatives, they must use their imagination, or they must use their experience to match patterns to the current situation [17]. There can also be situations where regression models do not do a good job capturing the environment when the criterion being modeled is not linear. Regression-based analyses can get around some nonlinearities by transforming information into cues that can be used in a linear model. However, this relies on the ability of the analyst to identify and apply this conversion and include it in the model, if such a transformation even exists. This plays directly into regression's second limitation, cue identification. Specifically, analysts must identify cues that can be linearly combined in a meaningful way so that the resulting model actually reflects the criterion being captured as well as the judgment strategy employed by the participants. This can be the most challenging part of regression-based JA.

B. Artificial Neural Networks

An alternative approach to modeling judgment that addresses some of the issues associated with regression-based JA comes from the field of ANNs. ANNs are models designed to generate judgments based on transformations of input information. The mechanism for these transformations is a collection of interconnected nodes designed to mimic the way neurons process information in animal central nervous systems. Specifically, each node has an adaptive weight that is used to modify the input information. Nodes are connected in multiple layers and node weights are adjusted progressively by a learning algorithm to fit the predicted judgment to the desired outcome. One of the most successful algorithms has been backpropagation (BP) [18], which enables the fitting of nonlinear judgment outputs.

Neural networks have advantages over regression-based JA techniques such as the lens model. First, they are capable of modeling nonlinear judgments [18], thus making them more flexible and potentially more accurate than lens models. Second, they do not require an analyst to explicitly identify the exact transformations of the available input data to produce linearly combinable cues. Because it does not require predictors (i.e., cues) to be linearly combinable, the neural net can transform the input data in whatever linear or nonlinear ways best achieve the desired judgment outcome. Thus, it is potentially easier to model human judgment using ANN techniques.

BP ANNs have been used to model human judgment in a number of different domains including automobile driver decision making, economics, and medical diagnosis [19]–[21]. They have also been used in a number of domains to model nonlinear environments [22]. However, these have focused on mimicking or simulating human judgments or environmental behavior rather than analyzing them. This is likely due to BP ANNs' black box nature: that an analyst cannot easily determine why a network is making the judgment predictions it is making, something that lens-model-based JA can do.

BP ANN analyses can be limited for several other reasons [22], [23]. First, network predictions can converge to local rather than global extrema. This can make predictions inaccurate. Second, the BP learning algorithm is very slow [24], which can limit when it can be used.

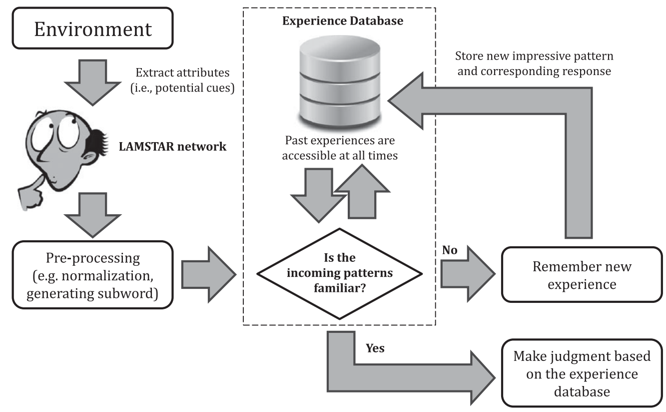

Fig. 2. LAMSTAR network fitting procedure.

C. Large Memory Storage and Retrieval Networks

LAMSTAR ANNs were developed to address some of the issues associated with BP networks [25]–[27]. LAMSTAR networks can handle larger quantities of data [28], are less sensitive to local extrema and thus more likely to converge to the desired outcome [25], [26], and are significantly faster than BP networks with comparable classification accuracy [29]. They can also work with a variety of different numerical data sources, with varying scales and accuracy between measures. Most importantly for this work is the fact that a LAMSTAR network is capable of providing information that allow analysts to evaluate how different independent variable inputs contribute to particular judgments. To understand this, one must understand how a LAMSTAR network learns (see Fig. 2).

Before a LAMSTAR network can be constructed, a preprocessing procedure must occur, where independent variables are categorized into subwords (collections of variables). Each variable can be assigned to only one subword (each variable is assumed to be quantitative). Then, during the network training process, any subword that significantly contributes to a judgment prediction is imprinted in a database as a neuron that can be used to produce judgments in the final model (see Fig. 3). Thus, over the course of the learning phase, each subword category will amass neurons in a self-organizing map indicating how many times a subword contributed to the development of a judgment. As such with a completed model, an analyst can evaluate how different independent variable subwords contributed to judgment prediction based on the total number of neurons (the link weight) associated with that subword. These totals can be used to compare how different words of variables contributed to different predictions, profile the judgment strategies of different judges, and compare judgment strategies. Details about the inner workings of the LAMSTAR network learning process can be found in [30].

Despite their advantages, LAMSTAR networks have never been used in JA.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

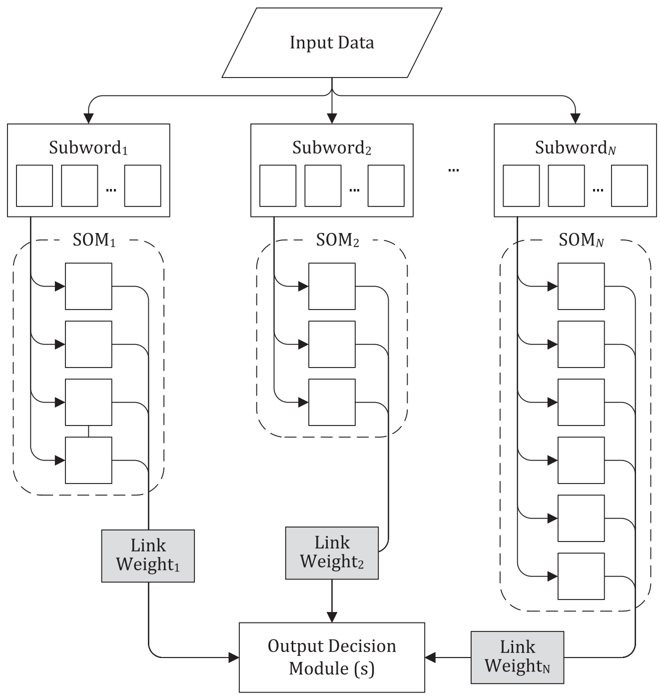

Fig. 3. Basic LAMSTAR architecture [30]. Note that SOM stands for self-organizing map.

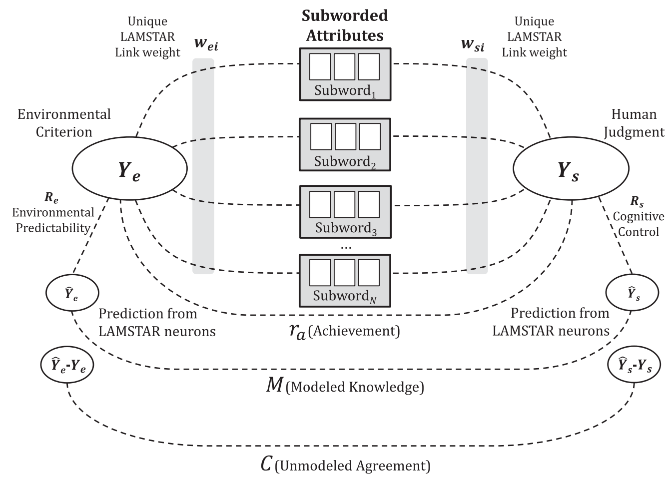

Fig. 4. Graphical representation of LAMSTAR-based JA. Note that i is an integer with range $[1, N]$.

D. Objective

In this work, we describe how LAMSTAR networks were used to develop a JA technique analogous to the double system lens model that avoids the linear assumptions and cue identification issues of regression-based judgment analyses. First, we outline how our LAMSTAR network JA process was formulated. Then, we use it to analyze data from an air traffic control conflict prediction task [12] and compare its results to those of a lens model analysis. Finally, we discuss our results and outline avenues of future research.

II. Large-Memory-Storage-and-Retrieval-Based Judgment Analysis

Our LAMSTAR-based JA is visualized in Fig. 4. In this reformulation, it is assumed that the analysts has a set of independent variables representing information from the environment that a human judge can synthesize into judgments. The analyst is also presumed to have the set of human judgments associated with values of those independent variables (Y_s) as well as an environmental criterion (Y_e) that represents the true value of the quantity being judged. The analyst must then manually separate the variables into groups representing general concepts from the environment that relate to the criterion. These should be both ecological and relate to concepts the human would use to make judgments. These groups are treated as the LAMSTAR subwords.

After grouping the independent variables into subwords, the LAMSTAR learning process can be performed over a given dataset. This will produce LAMSTAR network models for both the environment ($\hat{Y}_e$) and the human judgment ($\hat{Y}_s$) with estimated unique LAMSTAR link weights (w_{ei} and w_{si}) serving as the analogs of the ecological validities (r_{ei}) and the cue utilizations (r_{si}) from the double-system lens model.

From these data, additional parameters can also be calculated to help analysts understand the achievement of the human judgment that are analogs of the double-system lens model statistics. An achievement score r_a is the correlation between the criterion and the judgment (as with the lens model). We can also calculate environment predictability R_e and cognitive control R_s parameters as the correlation between the actual observations and those predicted by the models for the criterion and human judgment, respectively. A parameter M is used to represent a measure of the shared knowledge captured by the models as the correlation of $\hat{Y}_e$ and $\hat{Y}_s$ (note that this is the analog of G from the regression-based double-system lens model). Finally, unmodeled agreement (C) is the correlation of the residuals of each fitted model. It is important to note that because this new lens model formulation does not use multiple linear regression, these statistics will no longer relate to each other via the lens model equation (1).

III. Methods

A. Experimental Task

To evaluate our LAMSTAR-based JA technique and compare it with a traditional double-system lens model JA, we used data from a simple air traffic control task originally described in [12]. In this, participants were presented with a cockpit display concept representing traffic information (see Fig. 5). It is the task of the participant to monitor the relative trajectories of an intruding aircraft to ownship, represented in the center of the display. For each trial, the trajectory of the intruding aircraft was shown for a randomly selected time, uniformly distributed between 15 and 30 s. After this interval, the participant was tasked with judging the probability that the intruding aircraft would cause a loss of separation (get within 5 nmi of ownship) at its point of closest approach (POA).

B. Participants

In this study, we used data collected from six undergraduate engineering students, originally collected by Bass and Pritchett [12]. All participants were male. Note that this is a subset of the data collected for [12]. The reason for this is discussed under the next section.

C. Independent Variables and Experimental Design

In all trials, ownship is assumed to be navigating in straight and level cruise flight that is being controlled by the autopilot. The ownship

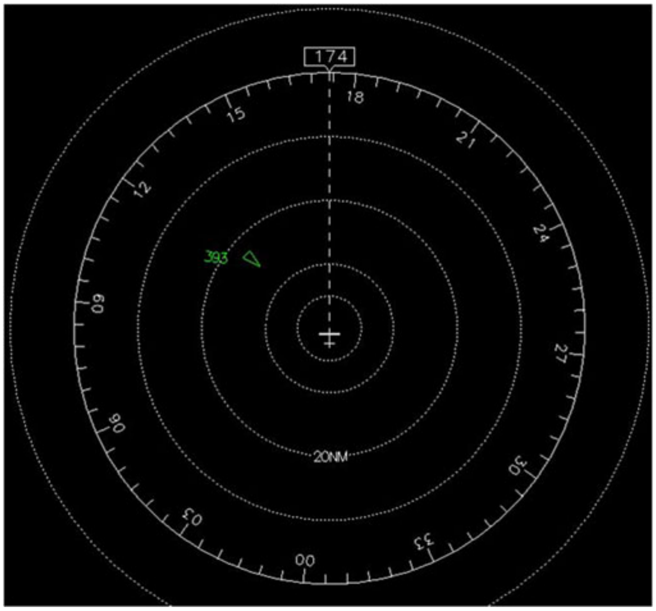

Fig. 5. Example scenario presented on the air traffic cockpit display concept. Ownship is the white airplane symbol at the center of the display. The incoming aircraft is the green triangle. Its heading in degrees is displayed next to it in green text. Six white concentric circles represent distance from ownship going from 5, 10, 20, 30, 40, and 50 nmi. A compass is represented on the 40-nmi circle.

retains a constant heading of 174°, an indicated airspeed of 400 kn, and altitude of 30 000 ft. In each trial, the intruding aircraft was assigned an altitude of 30 000 ft, one of six relative headings (±45°, ±90°, and ±135°), and five indicated airspeeds (300, 350, 400, 450, and 500 kn). The trajectory of the incoming aircraft was also determined by the distribution of error in the air speed, $\mathcal{N}(\mu = 0$ nmi/h, $\sigma = 15$ nmi/h); heading (*CourseError*), $\mathcal{N}(\mu = 0°, \sigma = 3°)$; and lateral position $\mathcal{N}(\mu = 0$ nmi, $\sigma = 0.27$ nmi). Finally, six lateral positions of the intruder aircraft were selected to ensure a uniform distribution in the probability of an LOA.

These parameters determined the independent variables for each trial, all of which were measured at the time human judgments were made. These included the speed and heading of ownship (*SpeedOwn* and *HdgOwn*) and the intruder (*SpeedIntr* and *HdgIntr*) and the error in these values (*SpeedError* and *CourseError*). It also resulted in human perceivable distances between the aircraft: the displayed distance between them (*RangeDisp*); ownship's distance to the PCA (*OwnDistToPCA*); and the intruders distance to the PCA (*IntDistToPCA*), all in nmi. Additionally, a derived variable representing the expected horizontal miss distance (the lateral separation at the PCA; *MDist*) was calculated from the aircraft trajectories and its associated error. A final derived variable (*Noise*) was calculated as "the signed reciprocal of the standard deviation of the horizontal position error" [12] to represent the position error and its relationship to the calculated probability of an LOA. The independent variables were used to create 45 unique representative trials. Each participant saw the same randomly ordered trials (see [12] for more details).

Bass and Pritchett [12] collected data from 16 participants over three phases with up to 180 trials per phase. There were 45 trials per session within a phase. We specifically used a subset of these data for several reasons. First, this work was not interested in exploring the relationship between humans and automation judgment assistance, which was the subject of the later phases in [12]. Thus, we only used

TABLE I
LAMSTAR Model Subwords

Subword	Independent Variables		
1	*SpeedError*	*SpeedOwn*	*SpeedIntr*
2	*CourseError*	*HdgOwn*	*HdgIntr*
3	*RangeDisp*	*OwnDistToPCA*	*IntDistToPCA*

data from the first phase (the "training" phase) because it collected human judgments based purely on environmental cues. Additionally, to facilitate direct comparison between models derived from participants, we wanted to use participants that saw the same trials at the same time in the experiment. Due to the counterbalancing in [12], this meant that only the trials of six participants from a single trial session could be compared. Thus, this limited our dataset to six participants with 45 trials per participant. The trials we used came from the first session of the participants' "training" phase.

Despite these reductions, the data should be sufficient for constructing the regression and LAMSTAR models we used in the presented analyses. Specifically, Bass and Pritchett [12] only used data from 45 trials to construct regression-based JA models for the last phase (the "prediction" phase) of their trials. Similarly, the number of trials (45) is larger than datasets used in the initial testing and validation efforts for LAMSTAR models (see [30]).

D. Dependent Measures

For each trial, a human judgment (Y_s) was collected representing what the participant thought the probability of a loss of separation was. Additionally, the actual probability of a loss of separation (Y_e) was calculated based on the relative positions and trajectories of the two aircraft and the represented speed, position, and course error.[1] Specifically, the aircraft's position was projected to the POA based on current values and errors in the aircraft's heading, position, and velocity. Based on the predicted POA, the distribution of miss distances was calculated. The probability of a loss of separation was then calculated using the cumulative distribution function of the horizontal miss distance. More details can be found in [12].

The independent variables and dependent measures were used to calculate the double-system lens model and LAMSTAR lens model parameters. For the lens model, multiple linear regression was used to fit linear models to Y_e and the Y_s values of each participant using *MDist* and *Noise* as the independent variables. These variables were used because they represented linearly combinable cues consistent with the judgment task and were employed in the original analyses by Bass and Pritchett [12]. These models were then used to calculate the remainder of the lens model parameters. For the LAMSTAR models, the nonderived independent variables were categorized into three different subwords (see Table I) each representing general categories of information an operator might use to make judgments: speed information, course information, and relative distance information, respectively. Note that we did not explicitly represent time to loss of separation in our subwords because it was not used in the actual calculation of the criterion and could not be explicitly observed on the display. It could, however, be derived using the information in the other subwords. The subwords were used to fit the model using LAMSTAR learning for the environment (Y_e) and the judgments of each participant (each participant's Y_s). Once the LAMSTAR model was created, it was used to calculate

[1]Speed, lateral position, and course error were all normally distributed with $\sigma = 15$ kn, $\sigma = 500$ m, and $\sigma = 3°$, respectively.

TABLE II
REGRESSION-BASED AND LAMSTAR-BASED LENS MODEL RESULTS

		Regression-based Lens Model				LAMSTAR-based Lens Model					
		$\hat{Y}_e$	R_e			w_{eSpeed}	$w_{eCourse}$	$w_{eDistance}$	R_e		
		$-0.073\ MDist - 0.179\ Noise + 0.878$	0.970			38	39	22	0.711		
Participant	r_a	$\hat{Y}_s$	R_s	G	C	w_{sSpeed}	$w_{sCourse}$	$w_{sDistance}$	R_s	M	C
1	0.274	$-0.004\ MDist - 0.073\ Noise + 0.519$	0.255	0.902	0.222	20	29	10	0.661	0.140	0.310
2	0.294	$-0.006\ MDist - 0.043\ Noise + 0.728$	0.328	0.962	-0.055	24	33	12	0.523	0.194	0.330
3	0.128	$0.002\ MDist - 0.035\ Noise + 0.615$	0.156	0.741	0.067	32	33	13	0.836	-0.103	0.235
4	0.128	$0.012\ MDist - 0.068\ Noise + 0.578$	0.318	0.349	0.090	22	21	14	0.623	0.023	0.221
5	0.245	$-0.001\ MDist - 0.068\ Noise + 0.600$	0.349	0.854	-0.198	17	23	7	0.553	0.158	0.240
6	0.346	$-0.011\ MDist - 0.024\ Noise + 0.656$	0.330	0.999	0.117	31	33	15	0.750	0.144	0.026
Average	0.238		0.290	0.935	0.041				0.674	0.093	0.229
SD	0.095		0.078	0.862	0.148				0.232	0.112	0.112
95% CI Low	0.163		0.231	0.552	-0.081				0.554	0.001	0.140
95% CI High	0.310		0.348	0.992	0.161				0.767	0.184	0.314

Note: SD stands for standard deviation. CI stands for confidence interval. All averages, SDs, and CIs are computed from correlation coefficients. Thus, each is computed by transforming each individual value using Fishers Z-Transformation, computing the statistic from the Z scores, and converting the result back into a correlation coefficient using the inverse of the Fishers Z-Transformation.

model predictions for all 45 trials, and the remaining LAMSTAR-based lens model parameters were calculated (see Section II).

E. Hypotheses and Statistical Analyses

Statistical tests were used to compare the lens model parameters of the two models. Specifically, Fisher's r-to-z' and z'-to-r transformations were used to compute the average, standard deviation, and 95% confidence interval bounds of each of the lens model parameters for each model over all of the participants. Because of dependence between the R_e values computed for the two models, a Steiger's Z statistic was used to compare them [31]. To compare the average G to M as well as the R_s and C values between models, we used two-tailed paired t-tests on the r-to-z' transformed correlations. All comparisons were conducted using a Šidák corrected 0.5 significance level of $\alpha = 1 - (1 - 0.05)^{1/4} = 0.0127$ to account for multiple comparisons.

Because the Lamstar model was expected to better capture the nonlinear and linear elements of both human judgments and the criterion, we hypothesized it would produce higher R_s and R_e values than the one based on multiple linear regression. No particular expectations were made about the relationships between G and M or between the two model's values of C.

IV. RESULTS

The lens model equations, subword weights, and lens model parameters from the analyses as well as the associated parameter averages are shown in Table II.

Achievement (r_a) across all participants and on average was low. This is consistent with the results reported by Bass and Pritchett [12]. The Steiger's Z test comparing the R_e values showed that it was significantly lower ($\bar{Z}_1^* = 6.98$, $df = 42$, $p < 0.01$) for the LAMSTAR model than for the regression-based model, a finding contradicting our hypothesis (this is discussed in Section V). The R_s values for each participant were always higher for the LAMSTAR model than for the regression-based one. The paired t-test of the r-to-z' transformed correlations produced evidence consistent with our hypothesis: that the R_s value was significantly higher ($t = 4.21$, $df = 5$, $p = 0.01$) for the LAMSTAR model than for the regression-based model.

While the G values from the regression-based models were always larger than the corresponding M values from the LAMSTAR models, there was no significant difference between them ($t = 3.13$, $df = 5$, $p = 0.03$). Similarly, although both the individual and average C values for the LAMSTAR models were larger than the corresponding individual and average C values from the lens model, there was no significant difference between the averages ($t = 2.35$, $df = 5$, $p = 0.07$).

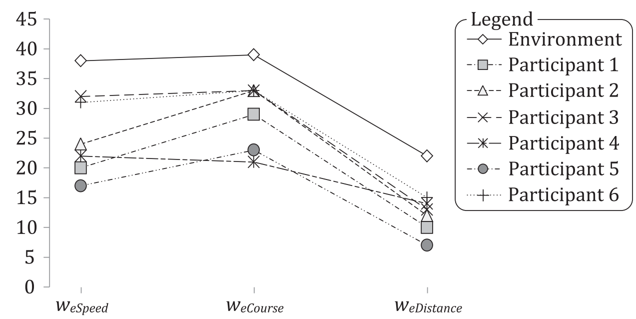

Fig. 6. Comparison of LAMSTAR link weights observed for each participant and the environment from Table II.

The results can also be used to compare the judgment strategies of participants to each other as well as the models generated for the environment. There are clear differences between the ecological validities of the environment model ($\hat{Y}_e$) and the cue utilizations of the subject models ($\hat{Y}_s$). Specifically, in all cases, the weight placed on *MDist* and *Noise* were nearly always an order of magnitude smaller than they were on the participant models. Similarly, two participants (3 and 4) actually weighted *MDist* positively, while all of the other participants (and the environment) weighted them negatively. Interestingly, participants 3 and 4 also had the lowest achievement scores.

The LAMSTAR link weights (w values) also provide insights into participant judgments (see Fig. 6). Specifically, with the exception of participant 4, all participants seemed to factor elements of course (w_{sCourse}) into their judgments more heavily than the other factors. In all cases, information related to distance ($w_{\mathrm{sDistance}}$) made the least contribution. Further, all of the participants seemed to make use of all three subwords less than the environment. There was no clear relationship between judgment strategy and achievement (r_a).

Finally, it is interesting to note that there is a much closer correspondence between R_e and R_s values in the LAMSTAR model than the regression-based one.

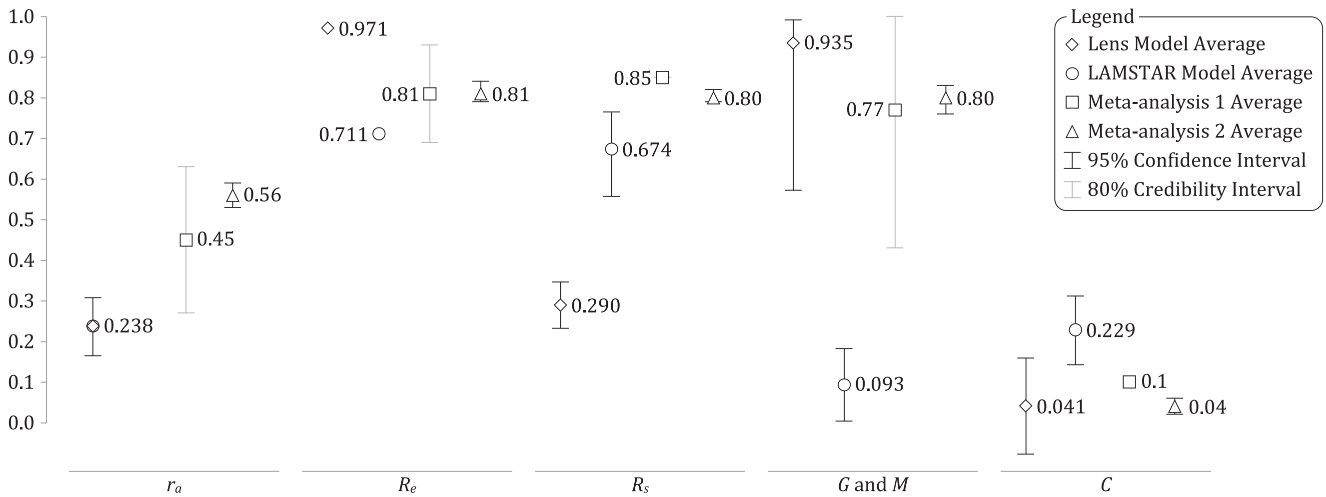

Fig. 7. Comparison of our lens model equation statistics (see Table II) with the averages from published lens model meta-analysis. Values are printed next to averages on the plot. Results from meta-analysis 1 are from [3] and [32]. Results from meta-analysis 2 are from [4]. Note that Kaufmann *et al.* [3] reported 80% credibility intervals; thus, these are reported above. A credibility interval was not reported with the C statistic [32], thus its absence in the figure above.

V. Discussion

The fact that the R_s values for the LAMSTAR models were significantly higher than those for judgment-based models was expected. This provides evidence that the LAMSTAR-based approach to judgment analysis is better at capturing human judgment strategies than the regression-based model, at least for the presented application. Meta-analyses have been conducted on regression-based lens model experiments across the academic literature [3], [4]. If we compare the results of our lens model analyses to the averages and confidence intervals of the meta-analyses (see Fig. 7), we see that cognitive control (R_s) appears to be lower in our data than in the meta-analyses. These results seem to show that the LAMSTAR approach could generally be more applicable in domains where human judgment is not well represented by a regression model.

While we hypothesized that the LAMSTAR model would better fit to the criterion, it is, *post-facto*, not shocking that this was not the case for the presented scenario. Specifically, the criterion values were generated based on the linear relationship shown in the fitted regression model. Thus, while the LAMSAR model did not fit the criterion data, here, it may do a better job of this for other task environments.

The comparison in Fig. 7 shows that our achievement values (r_a) were lower than average and that, for the regression-based model, both environmental predictability (R_e) and linear knowledge (G) were higher than the average. Our average unmodeled linear knowledge (C) was comparable to the meta-analysis average, although we did see a wide spread of values for C. Collectively, these results (along with those seen for R_s) suggest that the data used in our analysis are not representative of the data used by other researchers. Thus, the LAMSTAR model may only be applicable in a minority of situations. Future work should investigate what domains the LAMSTAR approach is appropriate for and why.

The analysis we used was based on the judgments of nonexperts. It is possible that the low R_s scores we observed for the regression-based models occurred because the novices used a less linear judgment strategy than would have been used by experts. Future work should investigate whether the LAMSTAR JA approach is appropriate for expert judges in the presented domain as well as others.

The LAMSTAR approach is convenient in that it can be used to predict human and environmental judgments within the range of possible inputs the same way the regression-based models can. Furthermore, the link weights provide some insight into what information factors into the predicted value. However, there does not appear to be an easily discernable relationship between the weights and achievement. Further, the link weights are not as easily interpreted as the cue utilizations produced for regression-based models. Future work will need to focus on how to best interpret LAMSTAR link weights.

Fig. 7 also gives us insights into how the LAMSTAR model parameters compare with those observed across multiple regression-based lens model analyses. While not quite statistically significant, the average value of M was lower than the average G value observed in this study. It was also lower than the G values seen in the meta-analyses. Given the nonlinear ways that LAMSTAR models are fit, it is not surprising that the M values were not only low, but lower than G values for regression-based models. This suggests that M may not be a useful measure when conducting a LAMSTAR-based JA. This should be investigated in future research.

The totality of our findings suggests that the LAMSTAR models and regression-based models are appropriate in different situations. Judgment analyses are predominantly concerned with understanding human judgments and comparing their judgment strategies to the environmental criterion [33], [34]. Thus, when linear models do a good job of capturing the human judgment and environment, the regression-based models will provide the most insights given the explanatory power of the ecological validities and cue utilizations. However, when a linear model does not fit the data well, the LAMSTAR model could be capable of providing models of and insights about the criterion and the human judgments, albeit without the deep explanative power of the regression-based model. Because JA can also be used as the basis for creating predictive models of human judgments, the LAMSTAR approach could definitely be advantageous in situations where human judgments do not appear to be linear.

This is the first study to investigate the use of LAMSTAR networks in JA. While the sample size used to fit the models was reasonable, because of restrictions in the utilized dataset, only data from six participants were considered. Thus, more work is needed. Future research should investigate how our results generalize for judgment tasks that both lend themselves to regression-based analyses and those that do not while considering larger numbers of participants and larger datasets in general.

Finally, there are extensions of JA that go beyond the double-system lens model. For example, the n-system lens model allows for analysis of and comparisons between multiple judges [2]. Future efforts could focus on using LAMSTAR networks for these types of lens model extensions.

VI. Conclusion

Given that this work is the first to adapt LAMSTAR networks for use in JA and that the approach appears to have utility, this work makes a significant contribution. However, this is just the first step toward understanding the implications of this novel approach. Future work should build on this infrastructure to extend the capabilities of this method to improve the ability of the method to provide insights into human judgments. Future work should also focus on using the method in other domains to better understand its generalizability.

Acknowledgment

The authors would like to thank Dr. E. J. Bass for letting them use data collected for [12] in the work presented here.

References

[1] E. Brunswik, *The Conceptual Framework of Psychology*, vol. 1. Chicago, IL, USA: Univ. Chicago Press, 1952.

[2] R. W. Cooksey, *Judgment Analysis: Theory, Methods, and Applications.* Waltham, MA, USA: Academic, 1996.

[3] E. Kaufmann, U.-D. Reips, and W. W. Wittmann, "A critical meta-analysis of lens model studies in human judgment and decision-making," *PLOS ONE*, vol. 8, no. 12, pp. 1–16, 2013.

[4] N. Karelaia and R. M. Hogarth, "Determinants of linear judgment: A meta-analysis of lens model studies," *Psychol. Bull.*, vol. 134, no. 3, pp. 404–426, 2008.

[5] C. J. Hursch, K. R. Hammond, and J. L. Hursch, "Some methodological considerations in multiple-cue probability studies," *Psychol. Rev.*, vol. 71, no. 1, pp. 42–60, 1964.

[6] L. R. Tucker, "A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd," *Psychol. Rev.*, vol. 71, no. 6, pp. 528–530, 1964.

[7] L. I. Dalgleish, "Decision making in child abuse cases: Applications of social judgment theory and signal detection theory," *Adv. Psychol.*, vol. 54, pp. 317–360, 1988.

[8] K. R. Hammond, *Human Judgement and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice.* Oxford, U.K.: Oxford Univ. Press, 1996.

[9] R. S. Wigton, "Applications of judgment analysis and cognitive feedback to medicine," *Adv. Psychol.*, vol. 54, pp. 227–245, 1988.

[10] T. R. Stewart, K. F. Heideman, W. R. Moninger, and P. Reagan-Cirincione, "Effects of improved information on the components of skill in weather forecasting," *Org. Behavior Human Decision Process.*, vol. 53, no. 2, pp. 107–134, 1992.

[11] R. W. Cooksey and P. Freebody, "Cue subset contributions in the hierarchical multivariate lens model: Judgments of children's reading achievement," *Org. Behavior Human Decision Process.*, vol. 39, no. 1, pp. 115–132, 1987.

[12] E. J. Bass and A. R. Pritchett, "Human-automated judge learning: A methodology for examining human interaction with information analysis automation," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 4, pp. 759–776, Jul. 2008.

[13] A. M. Bisantz and A. R. Pritchett, "Measuring the fit between human judgments and automated alerting algorithms: A study of collision detection," *Human Factors, J. Human Factors Ergonom. Soc.*, vol. 45, no. 2, pp. 266–280, 2003.

[14] E. J. Bass, L. A. Baumgart, and K. K. Shepley, "The effect of information analysis automation display content on human judgment performance in noisy environments," *J. Cogn. Eng. Decision Making*, vol. 7, no. 1, pp. 49–65, 2013.

[15] H. J. Einhorn, D. N. Kleinmuntz, and B. Kleinmuntz, "Linear regression and process-tracing models of judgment," *Psychol. Rev.*, vol. 86, no. 5, pp. 465–485, 1979.

[16] J. W. Payne, J. R. Bettman, and E. J. Johnson, *The Adaptive Decision Maker*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[17] R. M. Hogarth, *Educating Intuition*. Chicago, IL, USA: Univ. Chicago Press, 2001.

[18] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Representations by Back-Propagating Errors*. Cambridge, MA, USA: MIT Press, 1988.

[19] E. Fix, "Neural network based human performance modeling," in *Proc. IEEE Nat. Aerosp. Electron. Conf.*, 1990, pp. 1162–1165.

[20] P. Swicegood and J. A. Clark, "Off-site monitoring systems for predicting bank underperformance: A comparison of neural networks, discriminant analysis, and professional human judgment," *Intell. Syst. Accounting, Finance Manag.*, vol. 10, no. 3, pp. 169–186, 2001.

[21] H. H. Bruins, and R. W. Cooksey, "JANNET, a neural network approach to judgment analysis," in *Preradiation Dental Decisions in Patients with Head and Neck Cancer*. Utrecht, The Netherlands: Univ. Med. Center Utrecht, 2000. [Online]. Available: http://dspace.library.uu.nl/bitstream/handle/1874/393/c4.pdf

[22] D. Graupe, *Principles of Artificial Neural Networks*. Singapore: World Scientific, 2013.

[23] R. Rojas, *Neutral Networks: A Systematic Introduction*. Berlin, Germany: Springer, 1996.

[24] S. E. Fahlman, "An empirical study of learning speed in back-propagation networks," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-88-162, 1988.

[25] D. Graupe and H. Kordylewski, "A large memory storage and retrieval neural network for adaptive retrieval and diagnosis," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 8, no. 1, pp. 115–138, 1998.

[26] H. Kordylewski, D. Graupe, and K. Liu, "A novel large-memory neural network as an aid in medical diagnosis applications," *IEEE Trans. Inf. Technol. Biomed.*, vol. 5, no. 3, pp. 202–209, Sep. 2001.

[27] D. Graupe, "Large memory storage and retrieval (LAMSTAR) network," U.S. Patent 5 920 852A, Jul. 13, 1999.

[28] N. C. Schneider and D. Graupe, "A modified LAMSTAR neural network and its applications," *Int. J. Neural Syst.*, vol. 18, no. 4, pp. 331–337, 2008.

[29] J. M. Yoon, D. He, and B. Qiu, "Full ceramic bearing fault diagnosis using LAMSTAR neural network," in *Proc. IEEE Conf. Prognostics Health Manage.*, 2013, pp. 1–9.

[30] D. Graupe, "Large scale memory storage and retrieval (LAMSTAR) network," in *Principles of Artificial Neural Networks*. Singapore: World Scientific, 2013, pp. 203–274.

[31] J. H. Steiger, "Tests for comparing elements of a correlation matrix," *Psychol. Bull.*, vol. 87, no. 2, p. 245, 1980.

[32] E. Kaufmann, "Flesh on the bones: A critical meta-analytic perspective of achievement lens studies," Ph.D. dissertation, Soc. Beh. Sci., Univ. Mannheim, Mannheim, Germany, 2010.

[33] A. Kirlik and R. Strauss, "Situation awareness as judgment I: Statistical modeling and quantitative measurement," *Int. J. Ind. Ergonom.*, vol. 36, no. 5, pp. 463–474, 2006.

[34] R. Strauss and A. Kirlik, "Situation awareness as judgment II: Experimental demonstration," *Int. J. Ind. Ergonom.*, vol. 36, no. 5, pp. 475–484, 2006.