Coherence and correspondence in the psychological analysis of numerical predictions: How error-prone heuristics are replaced by ecologically valid heuristics

Yoav Ganzach[*]

Tel Aviv University

Abstract

Numerical predictions are of central interest for both coherence-based approaches to judgment and decisions — the Heuristic and Biases (HB) program in particular — and to correspondence-based approaches — Social Judgment Theory (SJT). In this paper I examine the way these two approaches study numerical predictions by reviewing papers that use Cue Probability Learning (CPL), the central experimental paradigm for studying numerical predictions in the SJT tradition, while attempting to look for heuristics and biases. The theme underlying this review is that both bias-prone heuristics and adaptive heuristics govern subjects' predictions in CPL. When they have little experience to guide them, subjects fall prey to relying on bias-prone natural heuristics, such as representativeness and anchoring and adjustment, which are the only prediction strategies available to them. But, as they acquire experience with the prediction task, these heuristics are abandoned and replaced by ecologically valid heuristics.

Keywords: numerical prediction, social judgment theory, cue probability learning, heuristics and biases.

1 Introduction

Numerical predictions — predictions in which a single, most appropriate numerical estimation of an outcome is required — are of central interest for both coherence-based and correspondence-based approaches to judgment and decisions. In the correspondence approach, numerical predictions are of central interest to Social Judgment Theory (SJT), derived from the Brunswikian ideas about probabilistic functionalism (Hammond, Stewart, Brehmer & Steinman, 1975). In the coherence approach, numerical predictions are of central interest to the Heuristic and Biases (HB) research program (e.g., Kahneman, Slovic & Tversky, 1982). The way such predictions are studied in these two traditions is different. In the SJT tradition they are studied by comparing them to measurable criteria in such a way that their correspondence with the environment can be evaluated. In particular, in the Cue Probability Learning (CPL) experimental paradigm, the most popular experimental paradigm in the SJT tradition, subjects learn the environmental relationship between cues (predictors) and outcomes, and are asked to generate predictions based on their learning.[1] Such a

paradigm allows for assessing the validity of the predictions by examining the correlation between the prediction and the true outcome (labeled the achievement index).[2] On the other hand, in the HB tradition, the validity of the predictions is assessed against normative prediction rules. For example, in their classic study, Kahneman and Tversky (1973) examined predictions of an outcome (GPA) from three predictors differing in their predictive validity, and found that they did not differ in their extremity, as measured by the prediction slope. Such predictions violate the basic normative (least square) law suggesting that predictions ought to be regressive: the higher the predictive validity of the predictor, the higher the extremity. Thus, at least at first glance it appears that the SJT approach to the study of numerical prediction is based on correspondence — it focuses on evaluating predictions based on their correspondence to an ecological criterion — whereas the HB approach is based on coherence — it focuses on evaluating predictions based on a normative standard.

One issue in evaluating coherence and correspondence approaches to judgment and decisions in general and nu-

---

[*]Address: The Leon Recanati Graduate School of Business Administration, Tel Aviv, 69978, Israel. E-mail: yoavgn@post.tau.ac.il.

[1]In the current paper we use the terms prediction and outcome rather than response and criterion, which are commonly used in the CPL literature. However, we use both predictor and cue, its equivalent in the CPL literature. We also use the term forecaster rather than judge, which

is commonly used in the CPL literature.

[2]Note that other paradigms that involve learning environmental relationship gained some prominence in the literature following the introduction of CPL. These paradigms study issues such as learning base rate utilization (Goodie & Fantino, 1996; Lovett, & Schunm 1999); utility maximization (Erev & Barron, 2005) or safety behaviors (Barkan, 2002).

merical prediction in particular is their reliance on different experimental paradigms. It is possible that research based on the correspondence approach relies on experimental paradigms such as CPL that tend to produce correspondence between judgments, decisions and the environment, whereas research based on the coherence approach relies on paradigms, such as the "one question" experiments (Kahneman, 2003), sometimes described as "experiments . . . conducted so that the word problems set up a 'trap' that subjects would fall into *if* they were using a particular heuristic" (Goldstein & Hogarth, 1997, p. 26), that tend to produce biases. Thus, the research programs associated with these two approaches take different views regarding people's adaptability to their environment. In particular, whereas SJT highlights people's adaptive behavior, and emphasizes accuracy and ecological rationality (e.g., Karelaia & Hogarth, 2008; Rieskamp & Otto, 2006), the HB program tends to emphasize biases, error and irrationality (e.g., Kahneman, Slovic & Tversky, 1982). In particular, in numerical predictions, CPL experiments almost always show that people are able to learn from experience and improve their predictions (see Brehmer & Joyce, 1988, and Cooksey, 1996, for reviews). On the other hand, Kahneman and Tversky (1973) are more pessimistic about people's ability to learn from experience and argue that: "Regression effects are all about us. In our experience, most outstanding fathers have somewhat disappointing sons, brilliant wives have duller husbands, the ill-adjusted tend to adjust and the fortunate are eventually stricken by ill luck. In spite of these encounters, people do not acquire a proper notion of regression . . . " (p. 249).[3].

In this paper I try to examine these two views by reviewing papers that have used a CPL experimental paradigm while attempting to look for heuristics and biases. The underlying theme is that both bias-prone heuristics and adaptive heuristics govern subjects' predictions in CPL. When they have little experience to rely on, subjects fall prey to natural, bias-prone, heuristics, primarily representativeness, but also anchoring and adjustment (Kahneman & Tversky, 1974). But, as subjects acquire experience with the prediction task, these heuristics are abandoned and replaced by adaptive, environmentally suitable, heuristics (e.g., Agnoli and Krantz, 1989; Nisbett, Krantz, Jepson & Kunda, 1983). Note that in contrast to the view that sees natural heuristics as adaptive (Gigerenzer & Goldstein, 1996, and Todd & Gigerenzer, 2007), the view that see them as error-prone (Kahneman & Tversky, 1974) is more in line with the data presented

in this paper, which suggest that biased heuristics are "natural" since they are associated with little experience.

Cue probability learning may be a most suitable experimental paradigm to study the interplay between natural and adaptive heuristics, since in the early phases of a CPL experiment subjects make predictions with no experience, whereas in the later phases they make predictions with abundant experience. Yet very little CPL research has examined this interplay. The reason, in my view, is that CPL researchers, adhering to the SJT approach, have not been interested in natural, error-prone heuristics, while HB researchers, de-emphasizing the effect of experience, have ignored cue probability learning, which emphasizes people adaptive behavior. Coming from the HB approach, I have also tended to ignore the central role of adaptive heuristics in intuitive prediction, and, despite reliance on CPL as a central experimental paradigm in my work (Czaczkes & Ganzach, 1996; Ganzach, 1993, 1994; Ganzach & Czaczkes, 1995; Ganzach & Krantz, 1990), have emphasized coherence rather than correspondence, heuristics of early rather than later phases, and extremity rather than achievement. In the current paper I attempt to examine this work again from the perspective of a dialectical distinction between coherence and correspondence.

1.1 Coherence, correspondence and precision: Measures of validity of numerical predictions

It is often argued that, whereas correspondence-based theories focus on ecological-validity, the match between decisions or judgments and environments[4], coherence-based theories focus on the validity of the rules underlying decisions or judgments, i.e., their consistency with normative mathematical models. However, this statement is appropriate only to the extent that the relationship between ecological validity and the mathematical models is ignored. That is, it could be argued that comparing intuitive decisions and judgments to normative models also reflects a concern with ecological validity, since these mathematical models were constructed to offer ecological validity.

In numerical prediction, particularly single-cue predictions, the normative model to which predictions are

---

[3]Kahneman & Tversky even argue that direct learning is also not likely to help. "As any teacher of statistics will attest, a proper notion of regression is extremely difficult to acquire" (ibid, p. 250). Or "Evidently, statistical training alone does not change fundamental intuitions about uncertainty" (ibid, p,251).

[4]I use the term ecological validity here to represent the validity of a system that quantitatively summarizes cues in order to predict an ecological criterion. This is a more general use of the term than the way it is commonly used as meaning the correlation between cue and criterion. (For example Cooksey, 1996, p.369 defines ecological validity as "Technically, and in Brunswick's original thinking, [ecological validity] is defined as the correlation between the values a particular cue takes on and the values a distal criterion takes on across a series of profiles containing that cue. Current usage has expanded such that the term is considered to be inclusive of any system for quantitatively summarizing cue emphasis in model predictions of an ecological criterion".

compared in the HB program is the regression model, which, by virtue of minimizing prediction error, namely the (squared) deviation between prediction and outcome, is an ecologically valid model. In this sense Kahneman and Tversky's (1973) experiment described above — even though it does not involve a direct comparison of intuitive predictions with outcomes — is still concerned with ecological validity of intuitive predictions.

Even though it is possible to cast the HB program as somewhat similar to SJT in that both are concerned with the ecological validity of the predictions, the two often differ in the yardstick they use to assess this ecological validity. In particular, in the SJT approach the yardstick for the validity of numerical predictions is achievement — the correlation between prediction and outcome — whereas in the HB approach the yardstick is appropriate slope. Note that both are desirable yardsticks for optimal predictions from an error minimization perspective, since they are both associated with reduction of prediction error. Yet, they reflect different emphases. In particular, forecasters interested in predictions that are ordered according to true scores — the prime feature of good prediction in tasks such as selection, for example — will be interested solely in the correlation between prediction and outcome. For this, forecaster achievement and not minimal error is the basis for an appropriate prediction model.[5] On the other hand, forecasters interested in prediction error — the prime feature of predictions regarding future values of financial instruments, for example — may be more interested in prediction extremity (see also Kahneman & Tversky, 1996, and Gigerenzer, 1996, for other domains in which disagreement about the appropriate criteria of judgment and decisions is at the heart of the debate between correspondence-based and coherence-based theories).

1.2 On the study of heuristics and biases in the CPL paradigm

Despite the major impact of Kahneman and Tversky's work on the way we view judgment and decision making in general and numerical prediction in particular, their ideas did not affect the way numerical predictions are studied in CPL. For example, the phenomena of excessively extreme predictions (overshooting), observed in a number of CPL experiments (e.g., Brehmer, 1973; Brehmer & Lindberg, 1970), received neither close scrutiny nor a theoretical explanation in CPL research.

Similarly, despite the power of CPL in studying the interaction between experience and reliance on heuristics, this experimental paradigm was rarely used by researchers in the HB tradition to study the effect of experience on heuristics. Thus, although there was some interest in the HB program regarding the effect of experience (e.g., Zukier & Pepitone, 1984; Nisbett, Krantz, Jepson & Kunda, 1983), this research was primarily a between-subject research, that is, it compared the output of experienced subjects to the output of subjects with little experience. On the other hand, the CPL experiments in the SJT tradition have inherently an important within-subject component, allowing for a continuous monitoring of the effect of experience on prediction strategies.

Below I describe studies that demonstrate how a better understanding of numerical predictions can be achieved by using ideas from both traditions. On the one hand, these studies demonstrate the insight into the processes underlying numerical predictions that is offered by the HB perspective on CPL. On the other hand, they demonstrate the understanding that can be gained from SJT-based studies regarding the processes underlying heuristics and biases, and the relationship between learning from experience and these processes.

2 Coherence and correspondence in a single cue probability learning

2.1 The modeling of single-cue prediction in HB and SJT

I start by comparing the HB perspective to the SJT perspective in the way they model the most simple prediction task, the prediction of a single outcome based on a single predictor in which the relationship between predictor and outcome is — as known to the forecaster – positive linear. That is, the prediction task is to make a prediction, $\hat{Y}_s$, based on a predictor, X, when the environmental, or true, relation between the true outcome Y_e and X is given by

$$Y_e = a + bX + \varepsilon. \tag{1}$$

In the SJT approach predictions are described by

$$\hat{Y}_s = a' + b'X + \varepsilon', \tag{2}$$

where a' and b' are parameters describing the prediction and ε' is the prediction error.

The normative prediction is given by

$$\hat{Y}_e = a + bX. \tag{3}$$

After experience, though, a learning process occurs and the parameters of the prediction model approach their

---

[5]For example, if P is the prediction (e.g., the prediction of performance) and T is the outcome (e.g., true performance), a forecaster who is primarily interested in prediction that will order candidates according to their true performance, that is, predictions that will maximize the probability that if $P_i > P_j$ then $T_i > T_j$, where the subscripts indicate the targets of the prediction (e.g., job candidates).

environmental values. That is, predictions are error ridden but not biased.

On the other hand, the HB approach suggests that predictions are biased. For convenience, and without loss of generality, I will discuss the HB approach regarding single-cue predictions in terms of the standardized values of the predictor, outcome and prediction. In this representation, the environmental relationship is given by

$$Z_{Y_e} = rZ_X + \varepsilon, \tag{4}$$

where Z_{Y_e} is the standardized value of the outcome, Z_X is the standardized value of the predictor, ε a random error and r the correlation between predictor and outcome.

In the HB tradition, such predictions are described as being made by relying on the representative heuristic — people make predictions which maximize the similarity between the input (the predictor) and the output (prediction). Such similarity is maximized by matching the extremity of the prediction to the extremity of the predictor. That is if the predictor is large (small), the prediction will be as large (small) as the predictor. Mathematically, such predictions are described by

$$Z_{\hat{Y}_s} = Z_X, \tag{5}$$

where $Z_{\hat{Y}_s}$ is the standardized value of the prediction on the *outcome* distribution.

This prediction is inconsistent with the normative least square prediction, which is given by

$$Z_{\hat{Y}_e} = rZ_X. \tag{6}$$

That is, normative predictions are regressive — the position of a normative prediction on the distribution of the outcome is less extreme than the position of the predictor on its distribution ($r < 1$). On the other hand, predictions by representativeness are non-regressive. The prediction is less extreme than the predictor, the smaller the correlations between predictor and outcome, the lower the extremity.

In sum, the crucial difference between HB and SJT is that the former suggests a model for prediction strategy when forecasters have little experience with the environment — a natural extremity matching strategy — whereas the latter does not have a good model for prediction based on little experience but offers a model by which changes in prediction strategy occur.

Finally, note that the distinction between natural erroneous strategies and ecologically valid strategies shaped by experience has already been discussed in the study of numerical predictions. Brehmer (1980) argued that people are naturally prone to rely on a positive linear strategy and only after substantial exposure to outcome feedback do they change their strategy to fit the actual functional relationship in the environment. In the terminology of

our discussion here, this strategy could be viewed as a positive-linear natural heuristic shaped by experience to fit the environment.

2.2 How are natural heuristics replaced with ecologically valid heuristics in single cue probability learning

2.2.1 Changes in reliance on representativeness

In previous studies (Ganzach, 1993, 1994) I compared predictions in two conditions, a representativeness-enhancing condition and a control condition, that were identical to each other except that in the former (but not in the latter) either the predictor or the outcome (feedback) was presented in such a way that reliance on representativeness was enhanced. Examples of representativeness-enhancing conditions are predictions in which the predictor is represented as bar graphs or percentile scores (both enhance representativeness since they provide the forecaster with a natural frame of reference against which the extremity of the predictor can be assessed; see Ganzach, 1993), or predictions in which outcome information is given as deviation of the outcome from the prediction (since such representation does not supply clear feedback, representativeness cannot be easily abandoned; see Ganzach, 1994).

The main indicator for reliance on representativeness in these studies was prediction extremity, operationalized as prediction slope. Two findings strongly emerged from these studies with regard to representativeness. (Figure 1 presents an example of the pattern of the results from one of the experiments, Ganzach, 1994, Experiment 2). First, predictions in the representativeness-enhancing conditions were more extreme than in the control conditions. Second, aside from this main effect, there was also a learning effect in the representativeness-enhancing conditions. Whereas in the beginning of the experiments the prediction slope in these conditions was as high as or even higher than the matching slope, there was a moderation of the slope throughout the experiment, and it approached the normative slope from above (little or no moderation was observed in the control conditions). Taken together, these results suggest that, when they have little experience, people rely on the representativeness heuristic in making numerical predictions, but that they also learn from experience and adopt more ecologically valid strategies. Note that these results are not consistent with the view that reliance on natural heuristics cannot be modified by experience (e.g., Kahneman & Tversky, 1973).

The second indicator that was examined in these studies was the consistency of the predictions — the correlation between the actual predictions and the prediction derived from the forecaster's model. Consistency
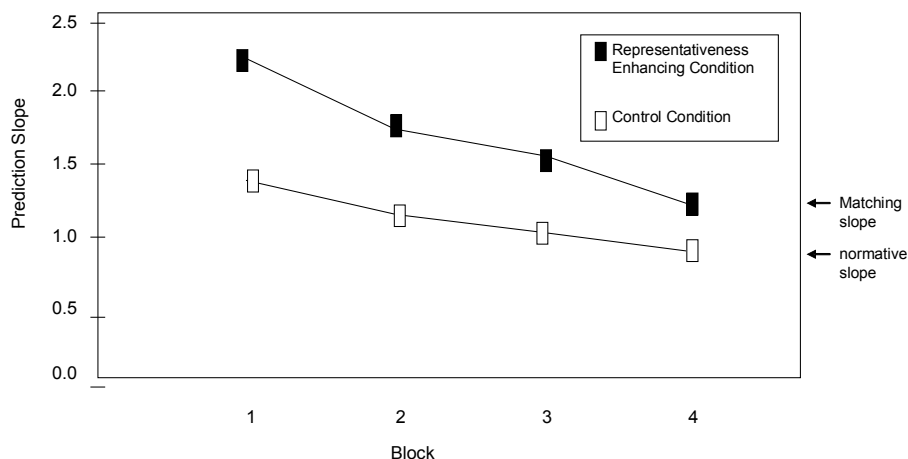
Figure 1: Mean prediction slope as a function of condition and 30 trials' block in Ganzach (1994), Experiment 2.

is also an indicator of reliance on representativeness reliance on matching should lead subjects to exhibit highly consistent predictions. In line with this, I found that in the representativeness-enhancing conditions predictions were more consistent than in the control conditions. However, there was almost no learning effect with regard to consistency: only small changes were observed in consistency throughout the experiment (but see footnote 4). Two factors are likely to hinder the learning of consistent predictions: The abandonment of representativness, which is essentially a linear prediction strategy; and the decrease in slope, which makes the retaining of linear prediction strategy more difficult.

2.2.2 Achievement, extremity, precision and representativeness

Being interested in indicators for reliance on representativeness, I analyzed the data in these single-cue probability learning (SCPL) experiments (Ganzach 1993, 1994) in terms of consistency and not in terms of achievement. But in SCPL, achievement is equal to consistency up to a multiplicative constant, the predictability of the environment — the correlation between the actual outcome and the outcome predicted from the environmental model.[6] Thus, in these experiments achievement can be directly inferred from consistency.

Since representativeness has a positive effect on both extremity and consistency, it may have contradictory effects on prediction validity as measured by extremity and

as measured by achievement. It leads to excessively extreme prediction, thus having a negative effect on prediction slope, the yardstick for validity in the HB approach. But it also leads to more consistent predictions, thus — since in an SCPL task consistency directly reflects achievement — having a positive effect on achievement, the yardstick for validity in the SJT approach.

These contradictory effects of representativeness on validity do indeed occur in the experiments discussed above (Ganzach, 1993, 1994). First, had validity been deduced from extremity, predictions in the control conditions would have been considered more valid, since they are closer to the normative slope, not displaying the excess extremity in the representativeness-enhancing conditions. But had validity been deduced from achievement, predictions in the representativeness-enhancing conditions would have been considered more valid, since they are more consistent. Second, had validity been deduced from extremity, predictions in the representativeness-enhancing conditions would have been considered more valid in the later phases of the experiments, since they are closer to the normative slope, not displaying the excess extremity of the earlier phases. But had validity been deduced from achievement, not much change in validity would have been detected throughout the experiments.[7]

To illustrate the implications of these contradictory effects of representativeness for the choice of an optimal prediction strategy, consider a forecaster who has weak

---

[6]In the lens model equation, achievement is the product of three indices: knowledge, environmental predictability and consistency. Environmental predictability is held constant in the experiment. Knowledge — the correlation between $\hat{Y}_s$ and $\hat{Y}_e$ — is equal to 1 in SCPL experiments, since both are linear functions of X (see Hogarth & Karelaia, 2007, and Lindell, 1976, for a discussion of the relationship between consistency and achievement).

[7]In some of the experiments we found a tendency of increasing consistency in the control conditions throughout the experiment, most likely as a result of the learning of the linear relationship between predictor and outcome. Although this was not a robust effect in our experiments, it has been clearly observed in other studies (see for example Dydycha & Naylor, 1966; Naylor & Clark, 1968). Thus, in the control condition there is a tendency of increasing validity throughout the experiment if validity is deduced from achievement, but no much change in validity if it is deduced from extremity.
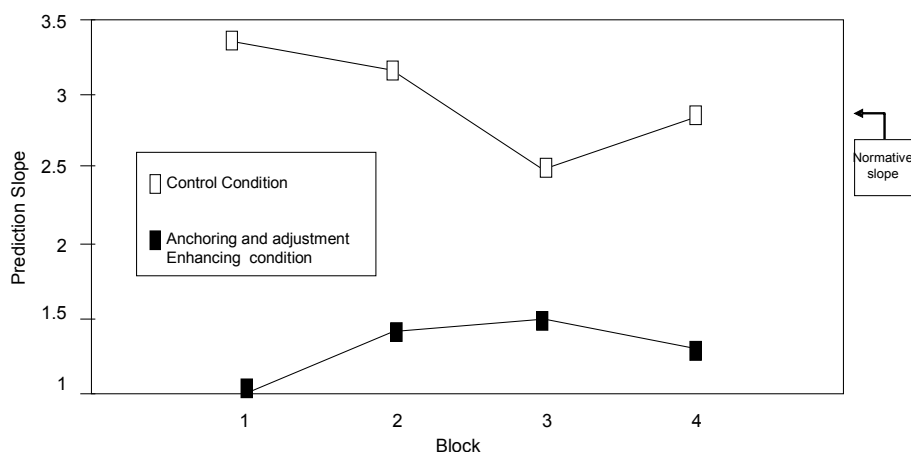
Figure 2: Mean prediction slope as a function of condition and 30 trials' block in Ganzach & Czaczkes (1996), Study 2B.

cognitive control (i.e., has difficulties generating the prediction implied by her — appropriately linear — prediction strategy; see Hammond et al., 1975), and makes predictions from a single low validity cue linearly related to the outcome. If the forecaster is interested primarily in achievement she may produce excessively extreme predictions (such that the deviations of her predictions from her model will be small relative to the variance of the predictions). This strategy, however, may increase prediction error. If the forecaster is interested primarily in prediction error, she may produce regressive and — given the inherent uncertainty about the appropriate slope — even forgo attempts to achieve an accurate slope and predict the mean of the outcome distribution for each value of the predictor. This may decrease prediction error but it also decreases achievement.

2.2.3 Changes in reliance on anchoring and adjustment

Consider predictions of an outcome distributed around zero (e.g. ranges between –26 and +26 with a mean of 0) as compared to an outcome with a distribution that differs only by an addition of a constant (e.g... ranges between 36 and 88 with a mean of 62). Since the mean of the former outcome is very salient, forecasters may use it as an initial value for their prediction and adjust on the basis of the extremity of the predictor. Since the adjustment is insufficient (Kahneman & Tversky, 1974), this may lead to over-regressive, excessively moderate, predictions.

Figure 2, taken from Czaczkes and Ganzach (1996, Study 2B) compares the prediction slope in an anchoring and adjustment enhancing condition to the prediction slope in its control condition (such a pattern has been replicated in other experiments (Czaczkes & Ganzach, 1996, Study 1, Study 2A, Study 3). It is clear from

this figure that (1) predictions in the anchoring and adjustment enhancing condition are less extreme than in the control conditions; and (2) there is a learning effect in which predictions become more (less) extreme in the anchoring and adjustment enhancing (control) condition, converging towards the optimal slope. Again, these results are not consistent with a strong view suggesting that reliance on natural heuristics is unlikely to be modified by experience (see above, Kahneman and Tversky, 1973, p. 249)

3   Coherence and correspondence in multiple-cue probability learning

So far I have not discussed the *integration* of a number of cues into a single prediction, an issue which is at the heart of probabilistic functionalism and SJT, and has been central to the CPL literature. In this section I review papers that used the Multiple Cue Probability Learning (MCPL) paradigm yet attempted to examine for heuristics and biases. In fact, I am aware of only three such papers, all of which used a two-cue probability learning task.

3.1   The linear integration of two cues

An early paper by Lichtenstein, Earle and Slovic (1975) offered a model for two cue predictions which are governed by representativeness. For simplicity I present a modified version of this model that includes only the features that are relevant to the current paper and discuss a prediction task with no error and orthogonal predictors. In such a prediction task, the representativeness heuristic suggests that the extremity of the prediction is a weighted

average of the extremity of the predictors:

$$Z_{\hat{Y}_s} = b_1 Z_1 + b_2 Z_2 \qquad (7)$$

$$\text{where } b_1 + b_2 = 1. \qquad (8)$$

In other words, it is an averaging strategy (e.g., Anderson, 1965; Jagacinski, 1995) in which the extremity of the two predictors is averaged to determine the extremity of the prediction. This strategy is non-normative since the normative (least square) model suggests that $b_1^2 + b_2^2 = 1$. (where b_1 and b_2 are, respectively, the correlations between the first and second predictor and the outcome).

In this two-predictor case, representativeness leads to excessively moderate predictions when there is no error-variance or when error-variance is low. As an example, consider two predictors with equal validity whose z-scores are 1 and 2. Predictions by representativeness will lead subjects to choose a prediction whose z value is 1.5 on the outcome scale, the average extremity of the two predictors. On the other hand, the normative prediction in our example of equal validity, no error and orthogonal predictors has a Z value of $0.707 \cdot 1 + 0.707 \cdot 2 = 2.13$. Thus, in this example intuitive predictions are under-regressive. Hence, in this case, predictions may be associated with a natural heuristic that leads to excessively moderate predictions when little experience is available, and a learning process in which predictions become more extreme as experience with outcome feedback is accumulated. Note that, although the suggested effect of experience in this case is similar to the pattern observed in Figure 2 (i.e., an increase in extremity as a result of experience), the underlying heuristic is different: it is representativeness rather than anchoring and adjustment. Note also that in MCPL the gap between normative predictions and predictions by representativeness increases as the number of predictors increases. For example, in the case of four predictors, two with z-scores values of 2 and two with scores of 1, the gap between predictions by representativeness, which are again 1.5, and the normative predictions, which are now 3.0, is larger than the gap in the parallel two-predictor case. Thus, the process by which error-prone heuristics are replaced by ecologically valid heuristics may be more pronounced the larger the number of predictors.

In sum, in this section I presented Lichtenstein et al.'s (1975) model, which suggests that in MCPL, the prediction is chosen so that its extremity is the average of the extremity of the predictors — a strategy consistent with representativeness — and suggested two hypotheses yet to be examined derived from this model. The first hypothesis is that predictions in a multiple-cue prediction task will exhibit under-regressiveness, and the second is that, the larger the number of predictors, the larger this under-regressiveness.

3.2   The configural integration of two cues

The second paper that used the MCPL paradigm yet attempted to examine for heuristics and biases did so in the context of the learning of configural prediction strategies. Consider a two-cue prediction environment in which the relation between predictor and outcome is given by:

$$Z_{\hat{Y}_e} = b_1 Z_1 + b_2 Z_2 + b_3 \, max(Z_1, Z_2), \qquad (9)$$

where $max(Z_1, Z_2) = Z_1$ if $Z_1 > Z_2$, and $max(Z_1, Z_2) = Z_2$ if $Z_2 > Z_1$; and assume for simplicity that the validities of the two cues are equal and that both are orthogonal and positively related to the predictor. In this environment, subjects have to learn not only the linear relationships between predictors and outcome, but also a configural relationship in which the predictors' weights depend on their relative value. When b_3 is positive the relationship is disjunctive — the higher predictor has a higher weight, and when it is negative the relationship is conjunctive — the *lower* predictor has a higher weight.

In this environment the natural configural strategy is disjunctive when the task involves predictions regarding people (it is associated with a positivity bias towards people) and conjunctive when the task involves predictions regarding non-human objects (it is associated with a negativity bias towards inanimate objects).

Consider the task of learning a *disjunctive* environment when the predictions involve an *inanimate* object. In this task, one has to abandon two natural prediction strategies in order to produce valid predictions. First, she needs to abandon the most natural cue combination strategy — linear combination (see equation 7) — and adopt a configural strategy; and second she needs to abandon the most natural configural strategy associated with predictions of *inanimate* objects — a conjunctive strategy — and adopt a disjunctive strategy. On the other hand, the learning of a *disjunctive* environment when predictions involve people requires abandoning only the linear combination strategy, and once this strategy is abandoned, the adoption of the configural (disjunctive) strategy is relatively easy.

By examining the linear weights (b_1 and b_2) and the configural weight (b_3) in a two-cue probability learning experiment, it is possible to study how people respond to environments described by equation 9. The results of such experiments indeed show that people start with a linear combination strategy but gradually shift to a configural strategy. This shift is facilitated (hindered) if there is fit (misfit) between the natural strategy and the configural aspects of the environment (Ganzach & Czaczkes, 1995).

To understand the nature of the configural strategy that subjects learn from experience, we used verbal protocols in which subjects were asked to describe their strategy (Ganzach & Czaczkes, 1995). Many of the responses

specifically stated that the lower (higher) predictor had more weight in the prediction, or that the two predictors were (only one predictor was) necessary for a higher prediction. These responses are indicative of an abstract rule of conjunctive (disjunctive) strategies.

3.3 SCPL and MCPL: Experience with multiple determination

In this sub-section I describe a paper that links the processes underlying single-cue prediction to the processes underlying multiple-cue prediction and relate them to the heuristics that lead to ecologically valid predictions. Consider a forecaster who has to make predictions based on a single cue, but has experience with multiple determination of an outcome, that is, experience in making predictions on the basis of two cues. As discussed above, the natural strategy in two-cue predictions is to choose a prediction in which the extremity of the prediction is the weighted average of the extremity of the predictors (equations 7 and 8). Coming to the single-cue prediction, the forecaster continues to use the same weighting scheme, but the unknown predictor is represented by its typical value. Since this typical value is usually the mean of the distribution (Kahneman & Tversky, 1982) — that is, $Z_2 = 0$ in equation 7 — equation 7 becomes:

$$Z_{\hat{Y}_s} = b_1 Z_1. \tag{10}$$

And since $b_1 = 1 - b_2 < 1$, the prediction is less extreme than the predictor, that is, it is regressive.

David Krantz and I examined the effect of experience with multiple determination in a number of studies. For example, in one study (Ganzach & Krantz, 1990, Exp. 4), the outcome was a linear function of two predictors. In the experimental condition subjects received either one predictor (in the odd trials) or two (in the even trials), whereas in the control condition they received only one predictor in *all* trials. Consistent with the effect of experience with multiple determination, the odd trials predictions of the experimental group were less extreme than those of the control group.

In one of the experiments (Ganzach & Krantz, 1990, Exp. 1), we also examined whether the effect of experience with multiple determination of one outcome (e.g., grade) affected the single-cue prediction of another outcome (e.g., salary), and found no generalization. That is, although experience with multiple determination led to regressive predictions of the first outcome (grade), it did not lead to regressive predictions of the second outcome (salary). This result is consistent with the idea that the prediction heuristic that is learned as a result of experience with multiple determination is not a general, abstract, prediction strategy. I further discuss this issue below.

4 Discussion

In this section I discuss two issues. First, I discuss the implications of the studies reviewed in the paper for the concept of valid predictions and the way they should be studied. And second, I discuss what can be learned from these studies about the processes by which biased heuristics are corrected and valid predictions are learned.

4.1 Standards of validity

It is often argued that coherence-based theories focus on evaluating the rationality of judgment, whereas correspondence-based theories focus on the predictive validity of the judge based on some ecological criteria (Dunwoody, 2009; Hammond, 1996). In this paper I argue that, at least as suggested by the SJT and HB traditions in the study of numerical predictions, both approaches could be viewed as concerned with ecological validity, the difference between them lying in the criteria they use: achievement in the SJT tradition and prediction extremity in the HB tradition.

This distinction suggests that CPL experiments should attend to the forecaster's view of prediction validity — accurate ordering of true scores or minimal prediction error — perhaps by clearly defining to subjects what constitutes good prediction in the experiment, or by rewarding them accordingly (e.g., for minimal error or for achievement). This is rarely done in CPL experiments, if at all. Nevertheless — since at least some of the data reviewed in this paper indicate a strong learning process regarding prediction slope and no learning regarding achievement (consistency) — it seems that for the subjects in these experiments, prediction extremity (and perhaps minimal prediction error) and not accurate ordering of predictions was the implicit criterion for prediction validity.

Are there other domains in which the differences between coherence and correspondence approaches could be conceptualized in terms of different standards of ecological validity? In my view, to some extent the answer is yes. Much of the debate between the HB program and the Fast and Frugal heuristics (FF) program (e.g., Gigerenzer, 2004) could be viewed as a debate about the criteria for valid judgment of likelihood: probability judgments or frequency judgments. Consider for example the debate on the FF program and the HB program about overconfidence (e.g., Gigerenzer, Hoffrage & Kleinbolting, 1991; Brenner, Koehler, Liberman & Tversky, 1996). Both programs examine likelihood judgments (the probability or frequency of correct responses) against an environmental criterion (the actual percentage of correct responses), and the debate between them concerns the question of whether probability or frequency judgment is the appropriate criterion for ecological validity.

4.2 Learning processes

The experiments described in the previous sections suggest that prediction biases associated with natural heuristics are corrected when they are incompatible with the environment. But how are they corrected? One possible explanation is that the natural strategies leading to biases are not replaced by other heuristics, but by exemplar-based strategies in which the retrieval of concrete, similar previous examples plays a role in correcting these biases (e.g., Juslin, Olsson, & Olsson, 2003; Karlsson, Juslin & Olsson, 2008).

Another possible explanation is that the natural prediction heuristics that underlie the biases are modified to produce more valid predictions after experience with outcome feedback is obtained. For example, it is possible that, in the representativeness-enhancing conditions in the SCPL experiments described above, the matching strategy is replaced by a modified matching strategy in which people still use the extremity of the predictor as an input, but produce a prediction that is somewhat less extreme (depending on predictor validity) than the predictor. Similarly, it is possible that in the anchoring and adjustment-enhancing conditions subjects still anchor at the mean, but learn through experience to allow for more adjustment on the basis of the extremity of the predictor.

While it remains to be seen if these two modified natural heuristics do indeed play a role in learning from experience in SCPL tasks, the named error heuristic in MCPL could clearly be viewed as a modified natural heuristic, since it is essentially a version of a two-cue representativeness heuristic in which the missing predictor is replaced by its typical value.

This discussion suggests that the ecologically valid heuristics that are developed during most of the CPL experiments described here are nothing like general abstract strategies, intuitive counterparts of the relevant statistical rules. This conclusion is consistent with Gigerenzer and Goldstein's (1996) description of the ecologically valid heuristics as ecologically rational rather than rational. Indeed, the failure of the subjects who gained experience with multiple determination to generalize the regressive prediction strategy they learned to predictions based on new predictors is consistent with the non-abstract, context-specific nature of these ecologically valid heuristics (see also Todd & Gigerenzer, 2007).

However, in learning configural environments subjects did seem to learn more abstract rules, at least as evidenced by their verbalization of the strategies they used. Perhaps, in this learning task there were alternative natural strategies (associated with positivity and negativity in judgments) that could compete with the prominent natural strategy of linear integration. Thus, consistent with previous research (e.g., Edgell, Harbison, Neace, Nahin-

sky & Lajoie, 2004; Lagnado, Newell, Kahan & Shanks, 2006; Rieskamp & Otto, 2006), the data presented in this review also suggest that what is learned from experience in CPL experiments may vary in the level of abstraction and in the scope of generalizability.

One obvious difference between the SJT tradition and the HB tradition in the study of numerical predictions is that the former uses experienced-based predictions whereas the later uses description-based predictions (Hertwig, Barron, Weber & Erev, 2004). Note, however, that this is not an important difference between the HB research program and FF research program, currently the most prominent correspondence-based research program. In particular, many of the studies in the FF program have relied on description-based procedures (see Gigerenzer, Todd & the ABC research Group, 1999). In fact, it could be argued that manipulated experience, such as the experience subjects gain in CPL experiments, is rather irrelevant to both the HB program and the FF program. For the HB program it is irrelevant because biased strategies are natural, and should not be affected by experience. For the FF program it is irrelevant because ecologically valid strategies are often viewed as exploiting "hard-wired … cognitive and motor processes" (Gigerenzer, 2004, p. 64), that frequently emerge even without experience; for example, our data suggest that, if anything, biased strategies are natural, but experience does modify them, leading to ecologically valid strategies.

It could be argued that the inconsistencies between the results described here and the FF perspective (i.e., the dominance of biased heuristic at the early stages of the experiments), as well as the inconsistencies with the HB perspective (the effect of learning from experience) could be explained by the fact that predictions in CPL experiments do not reflect real-world predictions. Proponent of the FF perspective may argue that CPL tasks are not representative of the environment in which evolutionary processes occur and produce ecologically valid strategies. Proponents of the HB perspective would argue that CPL experiments lack important characteristics of real-world predictions since in many real world situations "outcomes are commonly delayed and not attributable to a particular action … variability in the environment degrades the reliability of feedback … there is no information about the outcome would have been if another decision had been taken … and most important decisions are unique" (Tversky & Kahneman, 1973, p.198). However, a plausible view of the results is that the predictions in the CPL experiments described here do reflect prediction strategies that people use outside of the laboratory. When little experience is available, these strategies are error-prone natural heuristics. When experience with feedback is gained, these are ecologically valid heuristics, which may be modifications of natural heuristics.

References

Agnoli, F. & Krantz, D. H. (1989). Supressing natural heuristics by formal instructions: The cas of the conjunctive fallacy. *Cognitive Psychology, 21*, 515–550.

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology. 70*, 394–400.

Barkan, R. (2002). Using a signal detection safety model to simulate managerial expectations and supervisory feedback. *Organizational Behavior and Human Decision Processes. 89*, 1005–1031.

Brehmer, B. (1973). Single-cue probability learning as a function of the sign magnitude of the correlation between cue and criterion. *Organizational Behavior and Human Decision Processes*, 9, 377–395.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica, 45*, 223–241.

Brehmer, B., & Joyce, C. R. B. (Eds.) (1988). *Human judgment: The SJT view*. Oxford, UK: North-Holland.

Brehmer, B. & Lindberg, L. (1970). The relation between cue dependency and cue validity in single cue probability learning with scale cue and criterion. *Organizational Behavior and Human Decision Processes, 5*, 542–554.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). verconfidence in probability and frequency judgments. *Organizational Behavior and Human Decision Process. 65*, 212–219.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications.* San Diego, CA: Academic Press.

Czaczkes, B., & Ganzach, Y. (1996). The natural selection of prediction heuristics: Anchoring and adjustment vs. representativeness. *Journal of Behavioral Decision Making, 9*, 125–140.

Dunwoody, P. T. (2009). Theories of truth as assessment criteria in judgment and decision making. *Judgment and Decision Making, 4*, 116–125.

Dydycha L. W. & Naylor, J. C. (1966). Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Decision Processes, 1*, 110–128.

Edgell S.E., Harbison J.I., Neace W.P., Nahinsky, I.D., & Lajoie, A. S. (2004). What is learned from experience in a probabilistic environment. *Journal of Behavioral Decision Making, 17*, 213–229.

Erev, I. & Barron G. (2005). On Adaptation, Maximization, and Reinforcement Learning Among Cognitive Strategies. *Psychological Review. 112*, 912–931.

Ganzach, Y. (1993). Predictor representation and prediction strategies. *Organizational Behavior and Human Decision Processes, 56,* 190–212.

Ganzach, Y. (1994). Feedback representation and prediction strategies. O*rganizational Behavior and Human Decision Processes, 59*, 391–349.

Ganzach, Y., and Czaczkes, B. (1995). The learning of natural configural strategies. *Organizational Behavior and Human Decision Processes, 63*, 195–206.

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

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103*, 592–596.

Gigerenzer, G. (2004). *Fast and frugal heuristics: The tools of bounded rationality*. In D. J. Koehler and N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 62–88). Malden, MA: Blackwell Publishing.

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

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Gigerenzer, G., Todd, P.M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford Univ. Press.

Goldstein, W. M., & Hogarth, R. M. (1997), Judgment and decision research. In W. M. Goldstein & R. M. Hogarth (Eds.), *Research on judgment and decision making: Currents, connections, and controversies* (pp. 3–68). Cambridge, UK: Cambridge University Press.

Goodie, S. G. & Edmund, F. (1996). Learning to commit or avoid the base rate error. *Nature, 380*, 247–249.

Hammond K. R., (1996). Upon reflection. *Thinking & Reasoning. 2*, 239–248.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinman, D. O. (1975). Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), *Human Judgment and Decision Process*. New York: Academic Press.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*, 534–539.

Hogarth, R. M., & Karelaia N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review. 114*, 733–758.

Jagacinski, C. M. (1995). Distinguishing adding and averaging models in a personnel selection task: When missing information matters. *Organizational Behavior and Human Decision Processes, 61*, 1–15.

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. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.

Kahneman, D., Slovic & Tversky (1982). Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press: London.

Kahneman, D. & Tversky, A. (1973). The psychology of prediction. *Psychological Review, 80*, 237–251.

Kahneman, D., & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions *Psychological Review, 103*, 582–591.

Karelaia, N., & Hogarth, R.M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134***, 404–426**.

Karlsson, L, Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts? *Judgment and Decision Making, 3*, 244–260.

Lagnado, D.A., Newell, B. R., Kahan, S., & Shanks, D.R (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, 135, 162–183.

Lichtenstein, S., Earle, T. C., & Slovic, P. (1975). Cue utilization in a numerical prediction task. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 77–65.

Lindell, M. K. (1976). Cognitive and outcome feedback in multiple-cue probability learning tasks. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 739–745.

Lovett, M. C. & Schunm, C. D. (1999). Task Representations, Strategy Variability, and Base-Rate Neglect. J*ournal of Experimental Psychology: General, 128*, 107–130

Naylor J. C. & Clark, R. D. (1968)/ Intuitive inference in interval tasks as a function of the validity magnitude and sign. *Organizational Behavior and Human Decision Processes, 3*, 378–399.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339–363.

Rieskamp, J., & Otto, P.E. (2006). A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135*, 207–236.

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science, 16*, 167–171.

Tversky, A. & Kahneman, D. (1986). Rational choice and the framing of decisions. Journal of Business, 59, 251–284.

Zukier, H., & Pepitone, A. (1984). Social roles and strategies in prediction: Some determinants of the use of base-rate information. *Journal of Personality and Social Psychology, 47*, 349–360