

Effects of average uncertainty and trial-type frequency on choice response time: A hierarchical extension of Hick/Hyman Law

J. Toby Mordkoff¹

Published online: 10 March 2017
© Psychonomic Society, Inc. 2017

Abstract Hick/Hyman Law is the linear relationship between average uncertainty and mean response time across entire blocks of trials. While unequal trial-type frequencies within blocks can be used to manipulate average uncertainty, the current version of the law does not apply to or account for the differences in mean response time across the different trial types contained in a block. Other simple predictors of the effects of trial-type frequency also fail to produce satisfactory fits. In an attempt to resolve this limitation, the present work takes a hierarchical approach, first fitting the block-level data using average uncertainty (i.e., Hick/Hyman Law is given priority), then fitting the remaining trial-level differences using various versions of trial-type frequency. The model that employed the relative probability of occurrence as the second-layer predictor produced very strong fits, thereby extending Hick/Hyman Law to the level of trial types within blocks. The advantages and implications of this hierarchical model are briefly discussed.

Keywords Hick/Hyman Law · Mathematical models · Response time models

The information-processing approach to human performance takes its name, at least in part, from the discovery of a strong, linear relationship ($r^2 > .950$) between the amount of information being transmitted and mean response time (see, e.g., G. A.

The author thanks Cathleen Moore, Jan Wessel, and the reviewers for their comments, and Taharat Khan and Jason Fiala for help in collecting the data. This work was supported by the University of Iowa.

✉ J. Toby Mordkoff
jonathan-mordkoff@uiowa.edu

¹ University of Iowa, Iowa City, IA, USA

Miller, 1956). This phenomenon is now referred to as Hick's Law or Hick/Hyman Law (Hick, 1952; Hyman, 1953). When Shannon's (1948) method of quantifying information is used, the law is usually written as:

$$\text{mean RT} = a + b \cdot H, \quad (1)$$

where H is average uncertainty and both a and b are regression coefficients that are estimated from data. The value of H depends on the number and relative frequencies of the various possible trial types included within a block of trials:

$$H = \sum p(i) \log_2(1/p(i)), \quad (2)$$

where $p(i)$ is the probability of trial-type i and summation occurs across all of the trial types. When \log_2 is employed, the units of H are bits. (The choice of base has no effect on the quality of the fit to the data, but Base 2 has become the standard.) Note that the quantity $\log_2(1/p(i))$ is also known as the *specific uncertainty* or *surprisal value* for trial-type i , which is denoted $h(i)$, such that $H = \sum p(i) h(i)$ is an alternative version of Eq. 2 and the long-winded name for *average uncertainty* is “weighted mean surprisal value”.

Hyman (1953) presented three separate methods for manipulating average uncertainty. The first method, which became the most popular, is to manipulate the number of trial types, m , with each possible trial type being equally likely (i.e., all $p(i) = 1/m$). When this method is used, a short-cut formula for calculating average uncertainty is available: $H = \log_2 m$. The second method uses a constant number of trial types, but manipulates their relative frequencies (and requires the use of Eq. 2). In general, as the trial frequencies become more unbalanced (i.e., more unequal), the value of H is decreased. For example, four equally likely trial types have an H of 2.00 bits, but one trial type with a probability of .70 coupled with three other trial types with probabilities of .10

each has an H of only 1.36 bits. Thus, one can manipulate average uncertainty without changing the number of trial types included in each of the blocks. (This will be the method used in the present work.) The third method of manipulating average uncertainty also uses a constant number of trial types, each of which occurs equally often *overall*, but includes informative sequential dependencies to alter surprisal. For example, if the trials are selected at random, then a block with two trial types that occur equally often has an H of 1.00 bits. If a sequential dependency is added, however, such as raising the probability of a given trial type being repeated to .80 (and, therefore, lowering the probability of a trial-type switch across adjacent trials to only .20), then the value of H is reduced to 0.72 bits. Finally, it should be noted that these three methods of manipulating uncertainty may be combined. In fact, nearly all of Hyman's (1953) conditions using the second method also used the first, and nearly all of his conditions using the third method also used the first or second.

What is both impressive and important about Hyman's (1953) findings is how all three methods of manipulating average uncertainty produced the same pattern of results. Even more, because the same participants were tested using all three methods (and combinations thereof), the specific regression coefficients could be compared, within each subject, across methods, and these were always identical. In other words, regardless of how average uncertainty was manipulated, a given subject's data always produced the same values of intercept and slope for Hick/Hyman Law (i.e., values of a and b in Eq. 1), along with a very high value of r^2 .

In the decades following its discovery, Hick/Hyman Law was subjected to myriad tests and survived mostly unscathed (see, e.g., Proctor & Vu, 2006; Schneider & Anderson, 2011, for a recent review). This work also revealed several important relationships between the law and other phenomena. These include the finding that the slope, b , of Hick/Hyman Law depends on both practice (e.g., Mowbray & Rhoades, 1959) and stimulus-response compatibility (e.g., Teichner & Krebs, 1974) and can even be zero when ideomotor mappings are used (e.g., Leonard, 1959); a similar interaction between average uncertainty and Simon congruence (e.g., Stoffels, van der Molen, & Keuss, 1989); and the independence of uncertainty and repetition effects (e.g., Kornblum, 1967). More recently, Wifall, Hazeltine, and Mordkoff (2016) have presented a revised version of Hick/Hyman Law that includes separate values of b and H for stimulus and response uncertainty (for situations under which the number or frequency of the stimuli and responses are not confound), but even this should be viewed as an expansion of the law, as opposed to a falsification. However, one serious limitation of Hick/Hyman Law remains unresolved.

Trial-type frequency effects

The main limitation of Hick/Hyman Law—a weakness that has been known from the very beginning (see, e.g., Hyman, 1953, pp. 193–194)—is that it only applies to the means for entire blocks of trials. When fitting the law to a set of data, the mean of all response times (RTs) from each block is regressed onto the average uncertainty for the blocks (see Eq. 1). When the specific trial types within a block all have the same relative frequency, then it could be said that Hick/Hyman Law is also predicting the mean RTs for each of the specific trial types, as all should be equal. However, when the specific trial types within a block have different frequencies (as is true when Hyman's second method of manipulating uncertainty is used), then the mean RTs across trial types is not a strong linear function of their surprisal values (e.g., Crossman, 1953) or their simple probabilities of occurrence (e.g., Attneave, 1953). The lack of a quality fit across trial types with different frequencies is not evidence against Hick/Hyman Law, because the law is defined at the level of blocks. In fact, the same data that produce a nonlinear function at the trial-type level (with an r^2 below .800) can simultaneously produce a near-perfect straight line at the level of blocks (with an r^2 above .950; see Figs. 1 and 2 for a new example of this).

The lack of a linear relationship between mean response time for a given trial-type, $mRT(i)$, and its surprisal value, $h(i)$, may be somewhat surprising, especially given that the combination of Eqs. 1 and 2 implies that:

$$\sum p(i)mRT(i) = a + b \cdot \sum p(i)h(i), \quad (3)$$

where the left side of the equation has also been expanded to reflect that the mean RT for an entire block is, by definition, equal to the weighted mean of the mean RTs for each of the trial types contained in the block. This might appear to suggest that the mean RT for a given trial type should be directly predicted by its surprisal value using the same regression coefficients, a and b , as from Hick/Hyman Law, but this is known to be false. Hyman (1953), for example, found that

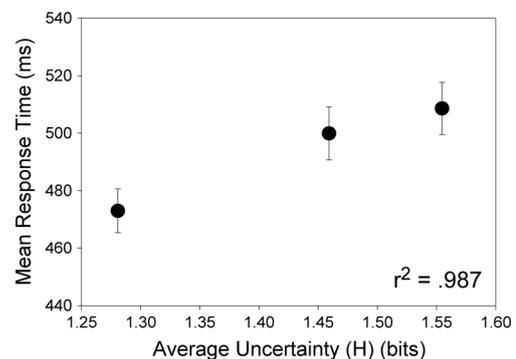


Fig. 1 Mean response time (with between-subject standard error) for each of the block types as a function of average uncertainty. (Note that within-subject error bars would be smaller than the plotted symbols)

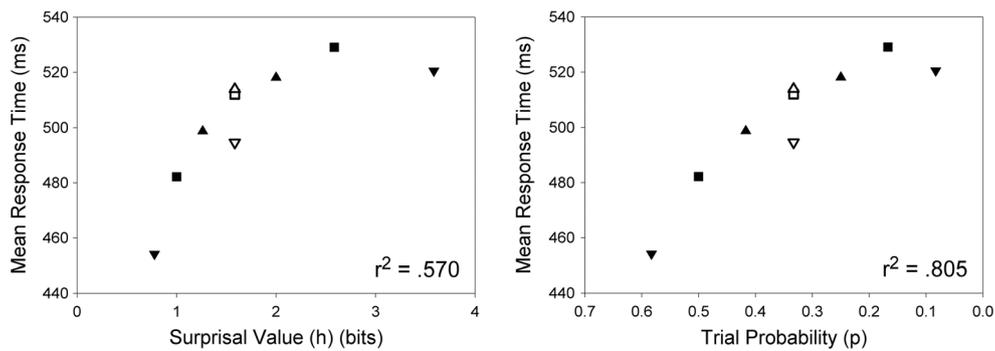


Fig. 2 Mean response time for each of the trial types as a function of surprisal value, $h(i)$ (left panel), and trial probability, $p(i)$ (right panel); note the inverted abscissa for trial-type probability. (\blacktriangle , \blacksquare , and \blacktriangledown indicate

the high-, medium-, and low- H block-level conditions; open symbols indicate the common-frequency conditions)

the value of b must be changed (by as much as a factor of 4) for the trial-type analysis to come close to correct, and subsequent work (e.g., Crossman, 1953; Kaufman & Lamb, 1966; Kornblum, 1967; Lamb & Kaufman, 1965) has made it clear that the trial-type fits are almost always nonlinear. In general, the fits for the specific trial types within a block are never as good as that for the blocks as a whole.

The response of Hyman (1953) to the inapplicability of Hick/Hyman Law at the level of blocks (which he called “conditions”) to the specific trial types within a block (which he referred to as “components”) was to write: “If we are interested only in the behavior of the condition means, the assumption of additive combination of the components will serve our purposes. If, however, we are interested in the behavior of components making up the conditions, we must find different laws and equations” (p. 194). Given the importance of understanding performance at the trial-type level, rather than only at the level of blocks, the goal of the present research was to find this new law or equation.

A hierarchical approach to Hick/Hyman Law

In defense of earlier work in this area, it’s worth repeating that Hick/Hyman Law was not designed to explain the performance of each of the specific trial types that are included within a block. The left side of Eq. 1 makes this clear: the law was only intended to explain differences in mean RT across entire blocks of trials. That Hyman and others have examined the fits to specific trial-types and found them to be poor does not change the success of the law in predicting performance at the level of blocks. However, attempts to model performance at both the block and trial-type level in terms of a single predictor, such as $h(i)$ or $p(i)$, have consistently failed (see, e.g., Fig 2), even when rather complicated approaches have been adopted (e.g., Kaufman, Lamb, & Walter, 1970).

In light of the inability of simple, one-predictor models to provide good fits at both the block and trial-type level, the current approach uses two predictors in a hierarchical model.

On the first pass through the data, differences in mean RT between entire blocks of trials are predicted by average uncertainty, H . This corresponds to the current version of Hick/Hyman Law (even with the amendment of Wifall et al., 2016). On the second pass, any remaining differences between the mean RTs for specific trial types within blocks are predicted by various instantiations of trial frequency (including two that are novel). This is the extension of Hick/Hyman Law.

To be considered successful, the extended model must pass at least two tests. First, the fit to the block-level data must remain at least as good as that provided by Hick/Hyman Law (i.e., the value of r^2 must remain extremely high, such as .950 or better). Second, the fit to the data from specific trial types must also be strong, and this must be achieved without reducing the fit to the block-level data. It was these two criteria that motivated the use of a hierarchical model. By having average uncertainty enter the regression first (and alone), the first layer of the model will always be the same as Hick/Hyman Law. By having the additional predictors enter second, after the regression coefficients for average uncertainty have been found and locked, the modeling of the specific trial types cannot alter the fit provided by Hick/Hyman Law at the level of blocks.

If more than one version of the extended model succeeds in passing these tests, then they should be compared or further explored. If no version of the model can pass these two tests, then the search for new laws and equations will have to continue.

Experiment

This experiment was designed under several constraints. First, at least three values of average uncertainty needed to be included, such that the high-quality fit provided by the current version of Hick/Hyman Law could be verified at the block level. Second, a wide variety of trial-type frequencies were also required, so the fits to the mean RTs for specific trial types could also be tested. Finally, one particular value of trial-type

frequency needed to be included in all of the blocks. This would provide a new and straightforward test of all simple models that attempt to explain performance only in terms of the frequencies of trials; any such model must predict no differences in mean RT across trial types with the same frequency of occurrence. As will be seen, the common-frequency conditions also provided the best data for discriminating between the various, competing, hierarchical models.

To achieve these goals while avoiding as many other issues as possible (e.g., the unwanted effects of an inconsistent mapping; see, e.g., Schneider & Shiffrin, 1977), the task for all blocks was three-alternative forced-choice, with the same mapping of stimuli to responses in each. In every block, one of the trial types occurred on one third of the trials. This was the common-frequency condition. Across blocks, the probabilities of the other two trial types were systematically varied (see Table 1; details below). This not only provided the range of trial-type frequencies required to test the hierarchical models but also created the three levels of average uncertainty required to verify Hick/Hyman Law.

Participants

A total of 72 students enrolled in Elementary Psychology participated, in partial fulfillment of a research-exposure requirement. Their ages ranged from 18 to 21 years, with a mean of 19.03 years. All but five of the participants expressed a preference to use their right hand. Fifty-one self-reported as female. The data from four participants (all right-handed, two females) were discarded prior to the analysis; two for using their cell phones while performing the task and two for switching hands for some of the blocks.

Apparatus, stimuli, and task

The experiment was conducted in small (2 m × 3 m) moderately lit rooms, using standard (Dell Optiplex) microcomputers running Windows-based (E-Prime) software.

Responses were made by pressing buttons on a small box (E-Prime Serial Response device) that was firmly attached to the table directly in front of the participant. All instructions, stimuli, and feedback were presented on a 15-in. diagonal (Dell) LCD screen, raised to eye level, approximately 80 cm in front of the participant.

The critical stimuli were orange, green, and purple filled squares that had previously been pilot tested for approximately equal luminance (via the flicker-fusion method). The squares were 1.25 × 1.25 cm, subtending a visual angle of 0.90° × 0.90° from the typical viewing distance. Each trial began with the appearance of a white fixation cross (0.70° × 0.70°) for 250 ms, followed by a blank screen for 250 ms, followed by the critical stimulus that remained visible until a response was made or 1,500 ms had elapsed. The intertrial interval was 2,500 ms.

Participants responded using the index, middle, and ring fingers of their preferred hand. The mapping of stimuli to responses was constant for each participant and completely counterbalanced across participants. Thus, for example, one sixth of all participants were asked to respond with their favored index finger if the stimulus was orange, with their favored middle finger if it was green, and their favored ring finger if it was purple. During training, feedback was provided after every trial, with more stress placed on accuracy than on speed, such that errors would be extremely rare and the analysis could focus on mean RT. During the experimental blocks, a tone was sounded (and the intertrial interval increased to five seconds) whenever an error was made; otherwise, only summary feedback at the end of each block was provided. Each block contained 36 trials (see below) and required less than 3 minutes to complete.

Design, procedure, and analysis

There were three different block types, as defined by their values of average uncertainty, as determined by the frequencies of the three types of trial (see Table 1). Each full block

Table 1 Average uncertainty, trial frequency, surprisal, probability, and summaries of performance for each of the block- and trial-level conditions

Block-level condition	Trial-level condition	$h(i)$	$p(i)$	Errors	Mean RT (all trials)	Mean RT (non-reps)
High H ▲ $H = 1.555$ (mean $p = .347$)	3 per 12	2.000	.250	0.5%	518 ms	532 ms
	4 per 12	1.585	.333	0.6%	514 ms	534 ms
	5 per 12	1.263	.417	0.6%	498 ms	533 ms
Medium H ■ $H = 1.459$ (mean $p = .389$)	2 per 12	2.585	.167	0.4%	529 ms	535 ms
	4 per 12	1.585	.333	0.4%	512 ms	534 ms
	6 per 12	1.000	.500	0.6%	482 ms	523 ms
Low H ▼ $H = 1.281$ (mean $p = .458$)	1 per 12	3.585	.083	1.0%	521 ms	521 ms
	4 per 12	1.585	.333	0.2%	495 ms	513 ms
	7 per 12	0.778	.583	0.4%	454 ms	500 ms

contained three subblocks of 12 trials each. (This was done to prevent the random selection of trial type from creating short-term values of frequency that deviated highly from what was intended. It also ensured that the first 12 trials in a block, which would be omitted from the analysis, always followed the planned set of trial-type frequencies.) The high- H condition used frequencies of three, four, and five trials per subblock of 12, producing an average uncertainty of 1.555 bits and surprisal values of 2.000, 1.585, and 1.263 bits, respectively. The medium- H condition used frequencies of two, four, and six per subblock of 12, with an average uncertainty of 1.459 bits and surprisals of 2.585, 1.585, and 1.000 bits, respectively. The low- H condition used one, four, and seven trials per subblock, with an H of 1.281 bits and $h(i)$ values of 3.585, 1.585, and 0.778 bits. Note how all three types of block have one trial type with a frequency of four per subblock, with a surprisal value of 1.585 bits and a probability of .333.

There are six possible ways to assign three different frequencies to three different trial types. When combined with three different sets of frequencies (i.e., the three different block types), this results in 18 different specific blocks, all of which were performed by every participant during the main part of the experiment. The order of the experimental blocks was pseudorandom, subject to the constraint that each set of three blocks included one from each of the three levels of block-level uncertainty (to equalize practice effects). At the beginning of each experimental block, the participant was shown the relative frequencies of the three trial types (in the form of percentages) beneath each of the colors for at least 2.5 seconds.

Prior to the experimental blocks, each participant was provided with detailed instructions and then six blocks of practice. During the practice blocks, all three trial types were always equally likely (i.e., the frequencies of each of the three trial types was always four per subblock of 12). The first practice block included written feedback after every trial; subsequent practice blocks only provided this feedback when an error was made. All blocks were followed by a summary display providing mean response time and percentage correct for the previous block. If accuracy was below 95%, a message was added, asking the participant to “please be more careful”.

As mentioned above, the first subblock (of 12 trials) of every full block was treated as “familiarization” with the new set of frequencies and, therefore, omitted from the analysis. Trials that immediately followed an error (less than 0.5%) were also omitted. As shown in Table 1, errors were extremely rare, such that no formal analysis of these data were performed.

Results

Before conducting the hierarchical analysis, a series of tests were performed to verify previous findings. The first of these concerned Hick/Hyman Law. To verify this, mean RT was

calculated for each type of block (ignoring the trial-types within the block). The findings from this analysis are shown in Fig. 1. As can be seen, the fit was remarkably strong, with an r^2 of .987. (The individual fits were also impressive, with a mean r^2 of .882 across participants).

The next question concerned the viability of models that attempt to explain the trial-level results in terms of a single predictor, either $h(i)$ or $p(i)$. One set of tests involved all of the data and corresponds to the left and right panels of Fig. 2. As can be seen (and consistent with previous findings; e.g., Kaufman et al., 1970), neither model provided an adequate fit, although $p(i)$ was approaching an acceptable level. The stronger test used only the data from the common-frequency conditions, because all simple models must predict that these three conditions would produce the same mean RT, as all would have the same predictor value, regardless of how trial frequency was being quantified. This prediction was easily disconfirmed (see the second, fifth, and eighth rows of Table 1 or the open symbols in either panel of Fig. 2): mean RT was still affected by block-level condition, even when trial-type frequency was exactly equated, $F(2, 134) = 16.58$, $p < .001$. In fact, the data from the common-frequency conditions were not only unequal but still conformed to Hick/Hyman Law ($r^2 = .931$), albeit with a much shallower slope than the analysis of all of the data.

With Hick/Hyman Law verified at the level of blocks and all simple models ruled out for trial-type frequency, it is time to consider hierarchical models. In all cases, the first layer of the hierarchy involved average uncertainty (only)—that is, Hick/Hyman Law was given priority. For each participant (separately), the block-level data were regressed onto average uncertainty (via Eq. 1). (This is the same subject-by-subject analysis that was mentioned above, with a mean r^2 of .882.) Next, the values of mean RT predicted by these regressions were subtracted from observed values to produce “residual times”—that is, the variations in mean RT that could not be explained by Hick/Hyman Law. This was done separately for all nine cells of the design, even though the regressions predict the same value of mean RT for all three levels of trial frequency for a given amount of average uncertainty. These residuals were then averaged across participants and then regressed upon four different possible predictors.

In the first pair of hierarchical models, the second-layer predictors were $h(i)$ or $p(i)$. The results from these analyses are shown in Fig. 3. As can be seen, while the hierarchical approach has raised the values of r^2 from what they were for the corresponding one-layer models, one critical problem remains: the mean residuals for the common-frequency conditions (see the open symbols in each of the plots) are still being predicted to be equal, as these models have no way of explaining the apparent interaction between average uncertainty and trial-type frequency.

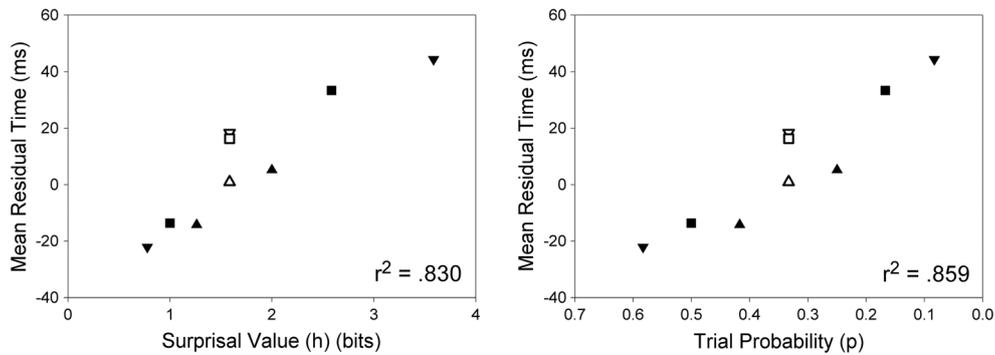


Fig. 3 Mean residual time for each of the trial types as a function of surprisal value, $h(i)$ (*left panel*), and trial probability, $p(i)$ (*right panel*); note the inverted abscissa for trial-type probability. (▲, ■, and ▼ indicate

the high-, medium-, and low- H block-level conditions; *open symbols* indicate the common-frequency conditions)

The last pair of models embrace the logic of hierarchical processing more closely and, thereby, correct the problems encountered by the first pair of models. Rather than using (raw) surprisal or trial-type probability as the second-layer predictors, these models used the deviation of these values from the mean for the block. Recall that an alternative label for *average uncertainty* is “weighted mean surprisal value.” Thus, when the data were processed for Hick/Hyman Law (on the first layer of the hierarchical analysis), the overall effect of average surprisal was being removed. In light of this, the new surprisal-based model used the difference between $h(i)$ and H as the second-layer predictor. This will be referred to as *relative surprisal*. Likewise, the new probability-based model used the difference between $p(i)$ and the mean value of $p(i)$ for the block, which will be referred to as *relative probability*.

The results for these last two models are shown in Fig. 4. Not only did these models provide better fits than either of the simple models (see Fig. 2) or the first pair of hierarchical models (see Fig. 3), but one achieved an r^2 of .942. When all four of the hierarchical models were directly compared, that using relative probability as the second-layer predictor was clearly the winner.

Besides producing the highest value of r^2 , this model also lost the least information (see, e.g., Akaike, 1973): The Akaike information criterion (AIC) value for the relative-probability model was 34.67, while those for raw h , raw p , and relative h were 44.46, 42.69, and 43.81, respectively. In addition, a series of stepwise regressions revealed that adding any of the other predictors to the relative-probability model never provided a significantly better fit to the data and always increased the value of AIC. In short, relative probability (alone) was the best second-layer predictor for the hierarchical model.

Even more, the relative-probability model had a second-layer intercept that was effectively zero (1.56 ms) and, therefore, lost no explanatory power when its intercept was constrained to be zero. (In fact, removing the intercept from this model increased the value of r^2 to .951.) This provides the additional advantage that the second layer of this model will cease to have any effect when all trial types have the same frequency of occurrence (i.e., when all $p(i) = \text{mean } p$). In other words, for those situations that do not need a revision or extension to Hick/Hyman Law (because all trial types are equally probable and, therefore, produce the same mean RT), this particular hierarchical model devolves to being identical to the existing version of Hick/Hyman Law.

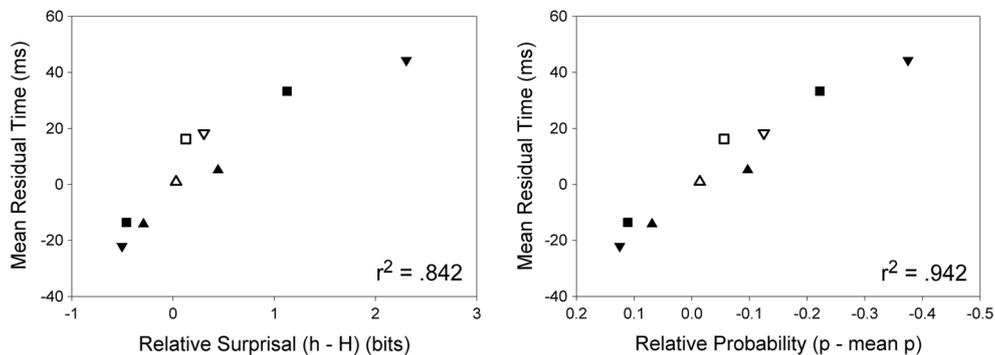


Fig. 4 Mean response time for each of the trial types as a function of relative surprisal value, $h(i) - H$ (*left panel*), and relative trial probability, $p(i) - \text{mean } p$ (*right panel*); note the inverted abscissa for relative

probability. (▲, ■, and ▼ indicate the high-, medium-, and low- H block-level conditions; *open symbols* indicate the common-frequency conditions)

Despite the advantage of the hierarchical model reducing to being identical to Hick/Hyman Law when all trial types within a block have equal frequencies, tests were also conducted on models under which average uncertainty and the second predictor entered the regression simultaneously, instead of one at a time in a specified order. As expected, in each case, the simultaneous model achieved a slightly better fit than the hierarchical model. For the model using relative surprisal, the value of r^2 increased from .842 to .881. For the model using relative probability, r^2 increased from .942 to .971. Note, however, the significant downside to taking this approach: The regression coefficients for average uncertainty (including the slope of Hick/Hyman Law) are no longer constrained to being the same across the various methods of manipulating uncertainty. In other words, to gain a few percentage in terms of total variance explained, the stunning convergence across methods that was demonstrated by Hyman (1953) could be lost. When one also notes the increase in computational complexity of the simultaneous model (i.e., the need for multiple regression, instead of simple regression), the hierarchical model is to be preferred.

Discussion

The goal of this work was to answer the challenge of Hyman (1953) and find a new law or equation that can explain the differences in mean RT that are observed across trial types with different frequencies. To be considered successful, the extended model needed to provide a strong fit to the data at the trial-type level while maintaining the existing strong fit at the block level. The hierarchical model that uses relative probability for the second-layer predictor meets both of these criteria by providing a very strong fit to the trial-level data while being identical to the existing version of Hick/Hyman Law at the level of blocks.

This hierarchical extension of Hick/Hyman Law to the level of trial types can be expressed as:

$$mRT(i) = a + b \cdot H + c \cdot \Delta p(i), \quad (4)$$

where $mRT(i)$ is the mean response time for a given trial type; a and b are the regression coefficients from the initial application of Hick/Hyman Law at the level of blocks (under which the different trial types within a block are collapsed); H is the average uncertainty for the block that included the trial type of interest; c is the slope from the second-layer regression of the response-time residuals (which is constrained to have an intercept of zero); and $\Delta p(i)$ is the deviation of the trial type's probability from the mean probability of all trial types within the same block (i.e., $p(i) - \text{mean } p$ for the block). As mentioned above, when all trial types within a block are equally probable, all values of $\Delta p(i)$ become zero and Eq. 4 reduces

to being the same as Eq. 1, which is the current version of Hick/Hyman Law. At the same time, in any situation under which the values of relative probability are not all zero, the hierarchical nature of the model still gives priority to the block-level effect.

One reasonable next question is whether the hierarchical extension of Hick/Hyman Law can be successfully applied to other, existing sets of data. Unfortunately, while many published experiments have included the needed conditions (starting with Hyman, 1953), very few have reported the data in sufficient detail for an analysis to be conducted. Of those studies that could be found with the needed response times, none have included three or more levels of average uncertainty coupled with three or more levels of trial frequency within each condition, and nearly all have used between-subject designs, which always include much more unexplained variance. With these limitations in mind, the model was still applied to two sets of previous data.

The first set of data comes from Kaufman and Lamb (1966), which used lines of varying lengths mapped onto vocal responses. The critical results are from the three “unequally likely alternatives” conditions, which were all two-forced-choice, but used trial-frequency splits of 90/10, 75/25, and 60/40. This produced three levels of average uncertainty ($H = 0.47, 0.81, \text{ and } 0.97$, respectively), albeit with only two levels of trial-type frequency in each case and no common-frequency condition. Despite this, the hierarchical model produced an r^2 of .913.

The second set of data is from Experiment 2 in Kaufman et al. (1970), which used the same general task as above but also varied the number of response alternatives between two, four, and eight. This greatly expanded the range of values for average uncertainty ($H = 0.54, 0.74, 0.90, 1.79, \text{ and } 2.40$), which helps to correct for the relative narrow range of values in both the present work and that of Kaufman and Lamb (1966). In this case, the hierarchical model produced an r^2 of .814, but it should be noted that the r^2 for the block-level effect was only .936, which is lower than usual, most likely because the manipulation of average uncertainty was applied between subjects, with only eight participants in each group. Future work should employ a broad range of average uncertainty (or a varied number of response alternatives), manipulated within-subjects, such the generality of the hierarchical model can be put to an even more rigorous test.

With regard to the wider literature, one implication of the present findings concerns the analysis of the results from experiments that only involve manipulations of within-block trial frequency (e.g., Irwin & Pachella, 1985; Mattes, Ulrich, & Miller, 2002; J. Miller & Hardzinski, 1981; Mordkoff & Yantis, 1991). In contrast to studies of block-level effects, where average uncertainty has been the predictor, these have often used the probabilities (or simple frequencies) of the various trial types as the predictor. The present work supports

this approach over those based on information theory's surprisal values, as long as all of the conditions are contained in one block. This holds because, within a block, differences in mean RT across trial types is explained (via Eq. 4) in terms of relative trial probability, and subtracting a common value—that is, mean $p(i)$ for the block—from all of the predictors would have no effect on the quality of fit. (It would merely shift the labels on the abscissa.) Conversely, however, when the manipulation of trial-type probability is done across blocks that have different values of average uncertainty (e.g., Gehring, Gratton, Coles, & Donchin, 1992; Hyman, 1953; Kaufman et al., 1970; Lamb & Kaufman, 1965; Maljkovic & Martini, 2005), then the prior regression of the mean RTs onto the values of H would be strongly suggested.

Another implication is much more theoretical (and somewhat speculative). The finding that the best fitting hierarchical model employs different metrics on each of the levels—that is, H on the first layer versus $\Delta p(i)$ on the second layer—suggests that the effects of block-level average uncertainty and trial-level relative probability arise in separate mechanism. Furthermore, average uncertainty was found to have the same effect on all trials within a block, regardless of the specific trial's frequency. Thus, the influence on performance of average uncertainty can be determined before stimulus onset, as it does not depend on the actual stimulus. In contrast (and by definition), relative probability does depend on the trial type, so the specific effect of this variable cannot occur until after the stimulus has been presented (and begins to be processed). All of this suggests that different mechanisms are likely responsible for each of these effects. Thus, the present extension of Hick/Hyman Law to the level of trial types implies that the further development of theoretical models (see, e.g., Erlhagen & Schöner, 2002; Schneider & Anderson, 2011; Wifall et al., 2016) might best be achieved by including at least two separate components: one that depends on average uncertainty and affects all trials equally, and another that depends on relative probability and affects specific trials differentially.

Some insight into the mechanistic differences between the block-level effects due to H and the trial-level effects due to $\Delta p(i)$ can be found in a post hoc analysis (suggested by a reviewer). One question that often arises whenever the frequencies of the different trial types are unequal is whether the observed effects are due to differences in how often specific trial types repeat across adjacent trials, as opposed to differences in overall frequency. This is usually addressed by reconducting the analysis with all repetitions excluded; if the pattern of results remains the same, then it could not have been caused by the repetitions. Using this logic, it has already been shown that differences in repetition rates are not responsible for block-level effects (i.e., Hick/Hyman Law; see, e.g., Kornblum, 1967; Wifall et al., 2016), and this finding was replicated by the present work: the block-level r^2 for all trials

was .987, while that for only the nonrepetitions was .946. In contrast, when the same reanalysis was applied to the trial-type data (i.e., the residuals from the block-level regression), the value of r^2 dropped from .942 to .083. This dramatic change occurred because there were often no differences in mean RT between the three different frequency conditions within a block when the repetition trials were omitted from the analysis (see the right-most column in Table 1). By the logic that is typically used, this implies that the trial-level effects come mostly from changes in the rates at which trials repeat, and not directly from differences in their overall frequencies. The dependence of the trial-level effects on differences in repetition rates, coupled with the independence of the block-level effects from the same, provides even stronger evidence for separate underlying mechanisms.

In summary, this work has successfully extended Hick/Hyman Law from the level of blocks to the level of trial types. When the various trial types within each block all have the same frequency, no change to the law is required, and no change is here made; differences in mean response time (across blocks) continue to be explained in terms average uncertainty, H . In contrast, when the different trial types within a block occur at different frequencies, the hierarchical extension of Hick/Hyman Law explains the residual differences in mean response time across trial types (within blocks) in terms of relative probability of occurrence, $\Delta p(i)$ (or relative probability of exact repetition). While it is true that the value of average uncertainty for a block depends on the probabilities of the trials within the block, the manner in which these two variables influence performance appears to be quite different, most likely arising in different mechanisms. Future work should explore this difference in detail.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Hkiado.
- Attneave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, 46, 81–86.
- Crossman, E. R. F. W. (1953). Entropy and choice time: The effect of frequency unbalance on choice-response. *Quarterly Journal of Experimental Psychology*, 5, 41–51.
- Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, 109, 545–572.
- Gehring, W. J., Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Probability effects on stimulus evaluation and response processes. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 198–216.
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4, 11–26.
- Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 30, 188–196.

- Irwin, D. E., & Pachella, R. G. (1985). Effects of stimulus probability and visual similarity on stimulus encoding. *American Journal of Psychology*, *98*, 85–100.
- Kaufman, H., & Lamb, J. (1966). Choice reaction and unequal stimulus frequencies in an absolute judgment situation. *Perception & Psychophysics*, *1*, 385–387.
- Kaufman, H., Lamb, J. C., & Walter, J. R. (1970). Prediction of choice reaction from information of individual stimuli. *Perception & Psychophysics*, *7*, 263–266.
- Kornblum, S. (1967). Choice reaction time for repetitions and non-repetitions: A re-examination of the information hypothesis. *Acta Psychologica*, *27*, 178–187.
- Lamb, J., & Kaufman, H. (1965). Information transmission with unequally likely alternatives. *Perceptual and Motor Skills*, *21*, 255–259.
- Leonard, J. A. (1959). Tactual choice reactions. *Quarterly Journal of Experimental Psychology*, *11*, 76–83.
- Maljkovic, V., & Martini, P. (2005). Implicit short-term memory and event frequency effects in visual search. *Vision Research*, *45*, 2831–2846.
- Mattes, S., Ulrich, R., & Miller, J. (2002). Response force in RT tasks: Isolating the effects of stimulus probability and response probability. *Visual Cognition*, *9*, 477–501.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
- Miller, J., & Hardzinski, M. (1981). Case specificity of the stimulus probability effect. *Memory & Cognition*, *9*, 205–216.
- Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 520–538.
- Mowbray, G. H., & Rhoades, M. V. (1959). On the reduction of choice reaction times with practice. *Quarterly Journal of Experimental Psychology*, *11*, 16–23.
- Proctor, R., & Vu, K.-P. (2006). The Cognitive Revolution at age 50: Has the promise of the human information-processing approach been fulfilled? *International Journal of Human-Computer Interaction*, *21*, 253–284.
- Schneider, D. W., & Anderson, J. (2011). A memory-based model of Hick's law. *Cognitive Psychology*, *62*, 193–222.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(379–423), 623–656.
- Stoffels, E.-J., van der Molen, M., & Keuss, P. J. G. (1989). An additive factors analysis of the effect(s) of location cues associated with auditory stimuli on stages of information processing. *Acta Psychologica*, *70*, 161–197.
- Teichner, W. H., & Krebs, M. J. (1974). Laws of visual choice reaction time. *Psychological Review*, *81*, 75–98.
- Wifall, T., Hazeltine, E., & Mordkoff, J. T. (2016). The roles of stimulus and response uncertainty in forced-choice performance: An amendment to Hick/Hyman law. *Psychological Research*, *80*, 555–565.