Research Article

THE INFORMATIONAL BASIS OF MOTION COHERENCE

David Gilden,[1] Eric Hiris,[2] and Randolph Blake[2]

[1]University of Texas at Austin and [2]Vanderbilt University

Abstract—*Observers judged the motion coherence of random-dot cinematograms Theoretical models were developed for coherence matches between cinematograms constructed from different angle distributions Evidence is presented that coherence matches are made on the basis of the Shannon-Wiener information entropy We show how the formal structure of information theory may be used to predict perceived pattern goodness when the underlying distributions of pattern alternatives are implicit in the judgment task*

Information theory has played two roles in the development of psychological theory As a mathematical tool for the quantification of sets of stimuli and responses, it has motivated normative models of human information processing (Attneave, 1959, Garner, 1953, Hake & Garner, 1951, Klemmer & Frick, 1953, Miller, 1956, Pomerantz & Kubovy, 1986) Historically, it also guided inquiry into the perception of pattern, with particular application to the Gestalt notions of pattern goodness and simplicity (Attneave, 1954, 1959, Garner, 1962, 1970, 1974, Garner & Clement, 1963, Pomerantz, 1977, Pomerantz & Kubovy, 1986) The hope was that the formalism of information theory might be used in descriptive models of perceptual organization In this regard, however, information theory never fulfilled its promise, primarily because of the intractability of associating individual patterns with the appropriate probability distributions that form the mathematical domain of information measurement In this article, we demonstrate that information measurement can be used effectively in descriptive models of perception when the statistical character of the stimulus pattern is made explicit in the discrimination task Specifically, evidence is presented that judgments of motion coherence are based on perceptual encoding of the Shannon-Wiener information entropy

The Shannon-Wiener information entropy is formalized (Shannon, 1948) as $H = -\Sigma p_i \log_2 p_i$, where p_i is the probability of the ith event and the sum runs over all possible events What the Shannon entropy measures is best understood as missing information, or uncertainty H is the number of yes/no questions that must be answered before there is no more uncertainty—when a particular event can be isolated from a set of possible events For a given set of N events $\{e_i\}$ with probabilities p_i, the uncertainty is maximized when the events are equiprobable In this case, the uncertainty is $\log_2 N$, and each binary decision provides the maximum of 1 bit of information

Early efforts to use H as a descriptive measure of pattern perception had to confront the problem that H operates on probability distributions defined over multiple events, and a single stimulus is just one thing Although H is defined in a theory having to do with information, and researchers tend to think of information as psychologically important, H is just a mapping defined on a set of event probabilities into the real numbers, and so is tacitly on the same footing as the mean and variance of a distribution Asking for the information of an individual stimulus event is as dubious as asking for its mean Garner's proposal that patterns are perceived in the context of inferred sets, and that such sets are generated by group operators such as reflection and rotation (Garner, 1962, 1970, 1974, Garner & Clement, 1963), provided the needed prescription for defining H on individual stimuli In this vein, good patterns became identified with those that were redundant under the group operators and so belonged to small inferred sets, sets in which a particular pattern could be identified with just a few bits of missing information Despite this insight into how a particular event could be embedded within a set of probable alternatives, the information formalism was never capitalized upon in modeling data In fact, the only numerical aspect of the inferred sets that was ever required was their size (see, e g , Garner & Clement, 1963) Thus, although information theory motivated theories of pattern goodness, the information entropy was never actually used in making predictions about how patterns would be perceived

Our experiments involve the perception of motion coherence Coherence and figural goodness are highly related concepts, and, in this sense, we follow Garner's lead in attempting to use information theory as a tool to clarify the content of Gestalt The key difference between our work and his is that we study systems that are themselves realizations of probability distributions Our stimuli are animation sequences consisting of individual elements whose motion paths continuously sample a distribution of directions By virtue of being inherently stochastic, this type of stimulus eliminates the need for an intervening theory to provide the probability setting that information theory requires The information entropy of a given animation sequence can be calculated exactly, and under these circumstances, it is possible to develop a quantitative model of information-based perception

PSYCHOPHYSICAL MEASURES OF MOTION COHERENCE

In our experiment, observers viewed random-dot cinematograms (RDCs), animation sequences in which the motions of individual dots are defined by a distribution of directions (Williams & Sekuler, 1984) Perceived coherence of motion in these stochastic animation sequences depends on the range of directions present in the distribution (Hiris & Blake, in press) We employed three different types of direction distributions, each defined by a single parameter (see Fig 1) Spike distributions (a proportion α moved straight down at $\theta = 0$, a proportion $1 - \alpha$ were distributed uniformly over 2π), Rectangular

Address correspondence to David Gilden, Department of Psychology, Mezes 330, University of Texas, Austin, TX 78712, e-mail gilden@psyvax psy utexas edu

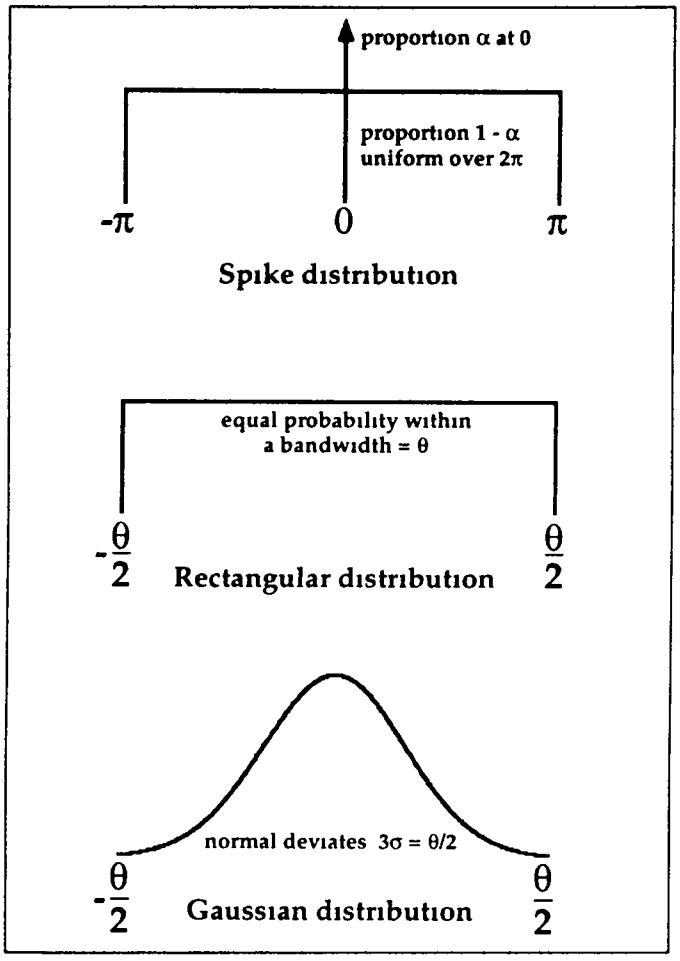

Fig 1 Three families of direction distributions Each family is indexed by a tunable parameter The Spike family of distributions is indexed by the proportion α of elements that move downward relative to the proportion $1 - \alpha$ that are distributed in angle uniformly over 2π The Rectangular family is indexed by the magnitude of the bandwidth θ determining the interval $[-\theta/2, \theta/2]$ within which all directions are equally probable The Gaussian family is also indexed by a bandwidth θ Direction distributions in the Gaussian family are centered at $\theta = 0$ on the interval $[-\theta/2, \theta/2]$, where the maximum angle that can appear in the display, $\pm\theta/2$, is defined to be at 3 standard deviations (i e , $\theta/2 = 3\sigma$)

distributions (all directions equiprobable between $\pm\ \theta/2$), and Gaussian distributions (directions normally distributed between $\pm\ \theta/2$, centered about $\theta = 0$) In this experiment, observers viewed two adjacent RDCs created from different distributions and judged which appeared to be more coherent

Methods

Our animation techniques are detailed elsewhere (Hiris & Blake, in press) In brief, RDCs were displayed on a 12-in monochrome monitor (66 7 Hz, P4 phosphor) controlled by a Macintosh II computer Viewed from a distance of 114 cm, each RDC consisted of 100 black dots (each 2 × 2 arc min) presented against a white background, with dots confined to a circular area 3 25° in diameter Between frames of these animation sequences, each dot chose a new direction from the range of possible directions specified by one of the three types of distributions For all RDCs, the average direction was constrained to be directly downward, and the speed of individual dots was always about 1°/s

On each trial, two RDCs were presented simultaneously side by side, 2 4° to the left and to the right of a fixation point While staring at this fixation point, an observer initiated a 2-s presentation of the two RDCs and during this time was allowed to look back and forth at the two RDCs Then, he or she indicated which RDC, left or right, appeared more coherent A 1-s rest period was enforced before the next trial could be initiated For a given block of trials, dot directions for one RDC were drawn from one type of distribution, and those for the other RDC were drawn from another type of distribution The position (left or right) of the two distribution types was randomized within a block of trials A block consisted of 210 trials comprising six repetitions of all combinations of the five standard RDCs drawn from the Spike distribution and seven comparison RDCs drawn from either the Rectangular or the Gaussian distributions Within a block of trials, the comparison stimuli were chosen from the same family of distributions, that is, the comparisons differed in terms of their bandwidth θ, but they were all Gaussian or all Rectangular distributed The five standard proportions were α = 4, 5, 6, 7, and 8 The seven comparison values of θ were chosen on the basis of pilot studies to bracket the points of coherence equality New standard and comparison RDCs were generated for each trial Eight observers, including one of the authors, completed five blocks of 210 trials, yielding 240 observations for a given combination of standard and comparison

Results

The raw data from this experiment consisted of the percentage of trials on which a given comparison Gaussian or Rectangular RDC was perceived to be more coherent than a given standard Spike RDC An example of a resulting psychometric function is shown in Figure 2 Probit analysis (Finney, 1971) was used to generate the best-fit curve for each psychometric function, and from each function we extracted the bandwidth (θ) associated with the point of subjective equality (PSE), that is, the comparison RDC judged to be equally coherent to the standard Spike RDC (see Fig 2) This analysis was applied to each standard α for each observer The median PSEs for the 8 observers are shown in Figure 3 as filled symbols The line connecting the PSEs for a given comparison family can be interpreted as an indifference function $\theta(\alpha)$ The Rectangular and Gaussian indifference functions form the basic units of analysis for theoretical modeling

THEORETICAL ANALYSIS OF COHERENCE EQUALITY

We have attempted to discover the properties of random-dot motion that observers use in comparing the coherence of RDCs This type of investigation has both formal and informal compo-

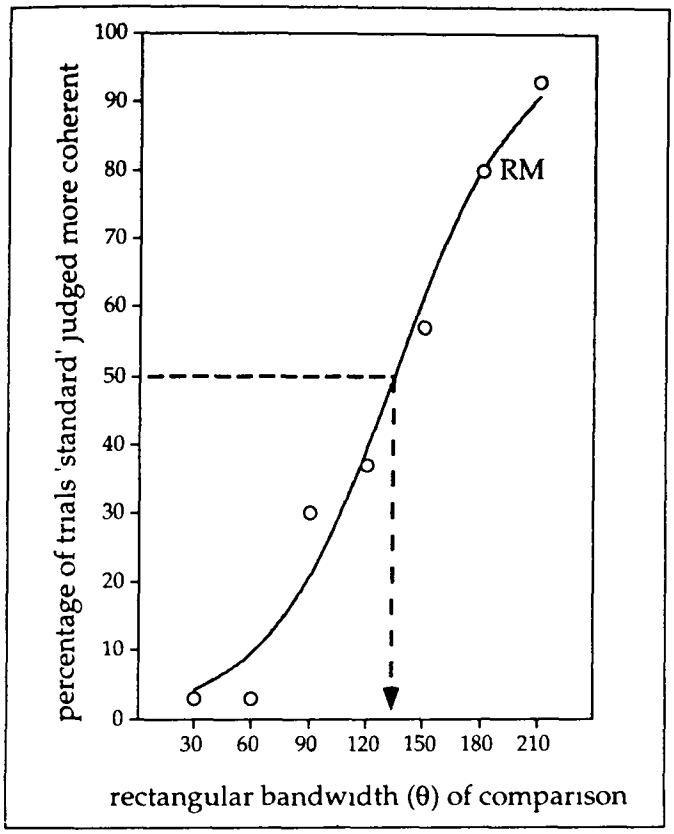

Fig 2 A representative psychometric function for standard Spike random-dot cinematograms (RDCs) with α = 6 paired with a set of seven comparison Rectangular RDCs Data from subject R M are shown Each data point is based on 30 observations, and the continuous line represents the best-fit curve derived from probit analysis The arrow shows the bandwidth θ at which the comparison and standard were judged to be equally coherent (i e , the 50% point on the psychometric function)

nents, and an understanding of our conclusions requires that we discuss the procedure in some detail At the outset, it must be recognized that the delineation of relevant features used in coherence judgment is necessarily informal There are an indefinite number of attributes of RDCs (the number of moments is infinite), and there is no external theory that can decide their relevance We have concentrated on four attributes that seemed potentially important on the basis of informal observations of the stimuli

Mathematical Modeling

Our goal is to develop a formalism that explains judgments of coherence Accordingly, we need some definition that links mathematical relations with observed relations in the data The following definition for Rectangular-Spike coherence matches is intended to make this link explicit

Let RDCs from the Spike distribution with proportion α be matched to RDCs from the Rectangular distribution with bandwidth θ Let F be a relation that maps direction distributions into the real numbers Then F explains a coherence judgment if $F(\alpha) = F(\theta)$

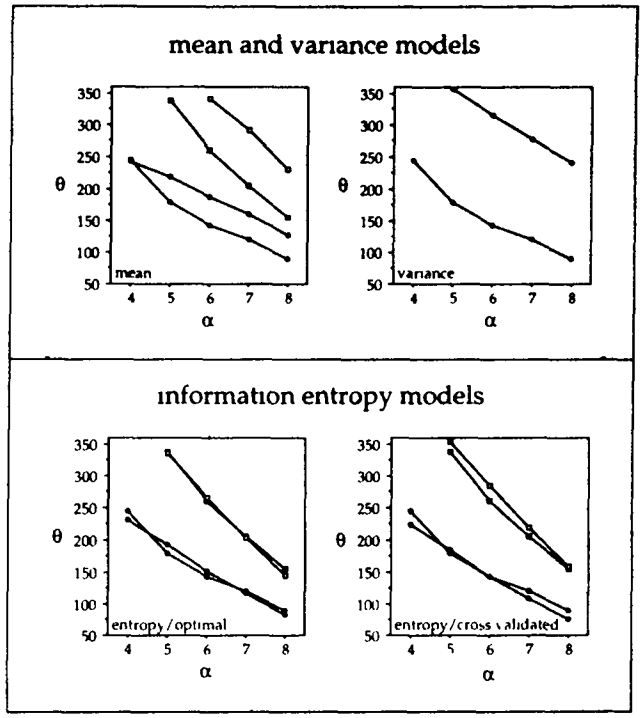

Fig 3 Observations of judged coherence equality (filled symbols) and models of these observations (unfilled symbols) The abscissa specifies different proportions (α) used as standards drawn from the Spike distribution, and the ordinate plots the bandwidths (θ) from the Rectangular and Gaussian distributions that yielded coherence matches The data are median coherence matches for 8 subjects In each graph (except that illustrating the variance model, in which the Spike and Gaussian distributions cannot be matched in the range of α studied), there are two sets of curves The upper curves (square symbols) depict data and models for Gaussian-Spike matches The 95% confidence intervals on Gaussian-Spike data were ±25° The lower curves (circular symbols) depict data and models for Rectangular-Spike matches The 95% confidence intervals on Rectangular-Spike data were ±13° The top two graphs illustrate models formulated in terms of the mean and variance of the downwards velocity component (cosine of dot heading) The bottom graphs show fits using the information entropy, $H = -\Sigma p_i \log_2 p_i$ The cross-validating curves are Gaussian-Spike predictions using a Rectangular-Spike optimization parameter and Rectangular-Spike predictions using a Gaussian-Spike optimization parameter As predictions, the cross-validating models have no degrees of freedom

A similar definition exists for Gaussian-Spike coherence matches If a relation F was able to explain our data perfectly, then we could solve for θ given α using the relation $F(\alpha) = F(\theta)$, and recover the observed indifference curve θ(α) In practice, a somewhat different procedure is required We take F and α as given, and find the θ′ that satisfies $F(\alpha) = F(\theta')$ [1] By inverting

1 Inverting the equation $F(\alpha) = F(\theta')$ cannot be done analytically for the types of relations that we consider Numerically, we accomplish the inversion by finding the root of the equation $G(\theta') = F(\alpha) - F(\theta')$ using Brent's method (Press, Flannery, Teukolsky, & Vetterling, 1986)

this latter equation, we define a theoretical indifference function $\theta'(\alpha)$ that may or may not bear a resemblance to the observed function $\theta(\alpha)$ In general, $\theta'(\alpha)$ is completely unrelated to $\theta(\alpha)$ Finding a relation F that creates an indifference function $\theta'(\alpha)$ that matches the data was the object of this investigation

Some of the relations, F, that we consider in this article require the specification of a free parameter, x In this case, we find the value of x that minimizes the difference between the theoretical and observed indifference curves The difference measure that we use is the sum of squared differences over the four or five standard values of α that define the observed indifference curve, $\theta(\alpha)$ Optimal values of the free parameter are determined separately for Rectangular-Spike matches and Gaussian-Spike matches Overall consistency is checked by cross-validation In a cross-validation procedure, the optimal parameter found for Rectangular-Spike matches is used to predict Gaussian-Spike matches, and vice versa These predictions have no degrees of freedom by virtue of the fact that the optimization was performed on different data using different stimuli

Definition of Models

The appearance of an RDC is dominated by two impressions First, it is evident that there is a general element drift in the downward direction Equally salient is the fact that there is disorder, not all of the elements move downward The attributes we have selected are different reflections of how the simultaneous presence of disorder and order might be perceived The attributes were as follows

1 The information entropy, H This relation computes the uncertainty of element motion from one moment of time to the next

2 The mean downwards component, μ_{down} In developing statistics defined on angles, proper measures of central tendency are defined on the components of position or velocity vectors (Batschelet, 1981) In our displays, there is no mean drift to the left or right, and the only nonvanishing mean component is along the vertical axis The appropriate measure of mean flow is

$$\mu_{down} = <\cos\theta>,$$

where <> denotes an ensemble average over elements, and $\theta = 0$ is straight down

3 The variance in the downwards component, defined as

$$\sigma^2_{down} = <\cos^2\theta> - \mu^2_{down}$$

4 A linear combination of the mean and variance known in the theory of extensive measurement (Krantz, Luce, Suppes, & Tversky, 1971) as risk Risk structures are defined only up to the specification of a free parameter that weights the contributions of the mean and variance Let r be a real number, $0 \leq r \leq 1$, then

$$\text{risk} = r\sigma^2_{down} - (1 - r)\mu_{down}$$

is the associated risk

Evaluation of Models of Coherence Perception

In evaluating the various models of coherence perception, we have had to recognize the fact that people do not discriminate directions of motion with arbitrarily fine resolution This is a particularly important issue in deriving the expectations of the entropy model, as greater resolution is associated with greater uncertainty For example, if people distinguished only between up and down, each motion element would introduce an uncertainty of at most a single bit We have attempted to make our calculations as realistic as possible by incorporating recent developments in the physiology and computational theory of angle discrimination

Current understandings of the underlying physiology of angle perception are formulated in terms of mechanisms that have Gaussian tuning curves in orientation Models of motion perception incorporate about 24 such mechanisms (Phillips & Wilson, 1984, Wilson, Ferrera, & Yo, 1992), although the exact number is somewhat dependent on the details of how the Gaussian is defined In the appendix, we describe how this formulation of the physiology is employed in a line element model (Watamaniuk, Sekuler, & Williams, 1989, Williams, Tweten, & Sekuler, 1991) The line element model allows us to create representations of RDCs in a hypothetical observer that is limited to N oriented mechanisms

Given that there is some uncertainty in the number of mechanisms that are implicated in the perception of motion direction, we have investigated the sensitivity of μ_{down}, σ^2_{down}, and H to N For $N \geq 12$, μ_{down} and σ^2_{down} are highly insensitive to the actual value of N We have evaluated these models using $N = 24$, which yielded results that were indistinguishable from $N = \infty$—infinite resolution in angle The information entropy model is generally more sensitive to the exact value of N, and we computed this model allowing N to range freely in order to find optimal fits to the matching data This freedom actually serves to provide an independent constraint on the acceptance of the entropy model If the optimal values of N had turned out to be inconsistent with current models of motion perception, this would have mandated rejection or reformulation of the entropy model In contrast, the free parameter in the risk model (r) that governs the weighting of the mean and variance is entirely unconstrained by both theory and experiment

The results of our calculations are shown as unfilled symbols in Figure 3 The mean and variance models do not explain the matching data The variance model is unable to account for any of the Gaussian-Spike matches because there are no values of θ for which $\sigma^2_{down}(\theta) = \sigma^2_{down}(\alpha)$ for any of the values of α employed in our experiment The risk model was no more successful than the mean model because optimal fits selected $r = 0$, that is, the best fits had no contribution from the variance Thus, the graph depicting the mean model also serves to depict the optimal risk model

The information entropy model was more successful The values of N mandated by the optimal fits were 24 and 28 for Rectangular and Gaussian matches, respectively When the optimal-fitting parameters are used in cross-validation, there is still a high degree of concordance between model and data Further validation of the entropy model comes from the observation that the derived values of the parameter, N, the number

of mechanisms used in angle discrimination, is in good agreement with the measurement made by Phillips and Wilson (1984) Convergent evidence that the values of N specified by the entropy model are meaningful comes from early measurements of channel capacity As described in the appendix, a line element model with 24 to 28 mechanisms produces about seven distinct categories or 2 8 bits of transmitted information This is exactly the number found by Pollack in measurements of channel capacity for angle categorization (summarized in Miller, 1956)

The models that have been considered in this study are certainly not exhaustive, nor could they be There is no sampling distribution of models, and we are not able to test the proposition that the information model is itself optimal We have included those models that seemed intuitively reasonable, and were initially surprised to find that the mean flow model performed so poorly Mean flow is a salient aspect of these cinematograms, and we initially conjectured that judgments of coherence would be made in terms of mean flow It was only after pilot data revealed the weakness of this model that we turned to considering information entropy and risk as bases for coherence matching We conclude from these models that information entropy appears to be encoded as a perceptual attribute of random motion stimuli and is used in judgments of coherence Coherence is implicitly a Gestalt property of shape, and this work is partial realization of attempts long since abandoned to provide a metric for the elusive and ill-defined notion of figural goodness

Acknowledgments—We would like to thank Jeff Schall for his insight into our data This research was supported in part by National Institutes of Health Grant EY07760 and Air Force Office of Scientific Research Grant 749620-93-1-0307

REFERENCES

Attneave, F (1954) Some informational aspects of visual perception *Psychological Review, 61*, 183–193

Attneave, F (1959) *Applications of information theory to psychology* New York Holt Rinehart & Winston

Batschelet, E (1981) *Circular statistics in biology* New York Academic Press

Finney, D J (1971) *Probit analysis* Cambridge, England Cambridge University Press

Garner, W R (1953) An informational analysis of absolute judgments of loudness *Journal of Experimental Psychology 46*, 373–380

Garner, W R (1962) *Uncertainty and structure as psychological concepts* New York Wiley

Garner W R (1970) Good patterns have few alternatives *American Scientist 58* 34–42

Garner W R (1974) *The processing of information and structure* Potomac, MD Erlbaum

Garner W R & Clement, D E (1963) Goodness of pattern and pattern uncertainty *Journal of Verbal Learning and Verbal Behavior 2*, 446–452

Hake H W & Garner W R (1951) The effect of presenting various numbers of discrete steps on scale reading accuracy *Journal of Experimental Psychology 42*, 358–366

Hiris E , & Blake, R (in press) Discrimination of coherent motion when local motion varies in speed and direction *Journal of Experimental Psychology Human Perception and Performance*

Klemmer E T & Frick F C (1953) Assimilation of information from dot and matrix patterns *Journal of Experimental Psychology 45* 15–19

Krantz D H , Luce R D Suppes, P & Tversky, A (1971) *Foundations of measurement Vol I Additive and polynomial representations* New York Academic Press

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

Phillips G C , & Wilson H R (1984) Orientation bandwidths of spatial mechanisms measured by masking *Journal of the Optical Society of America 1* 226–232

Pomerantz J R (1977) Pattern goodness and speed of encoding *Memory and Cognition, 5* 235–241

Pomerantz, J R & Kubovy, M (1986) Theoretical approaches to perceptual organization Simplicity and likelihood principles In K R Boff, L Kaufman & J P Thomas (Eds) *Handbook of perception and human performance* (pp 36 1–36 46) New York Wiley

Press W H Flannery, B P , Teukolsky, S A , & Vetterling, W T (1986) *Numerical recipes* Cambridge England Cambridge University Press

Shannon C E (1948) A mathematical theory of communication *Bell System Technical Journal 27* 379–423 623–656

Watamaniuk, S N J Sekuler R , & Williams, D W (1989) Direction perception in complex dynamic displays The integration of direction information *Vision Research 29* 47–59

Williams, D W , & Sekuler, R (1984) Coherent global motion percepts from stochastic local motions *Vision Research 24* 55–62

Williams, D W Tweten, S & Sekuler R (1991) Using metamers to explore motion perception *Vision Research 31* 275–286

Wilson H R Ferrera, V P & Yo, C (1992) A psychophysically motivated model for two-dimensional motion perception *Visual Neuroscience 9* 79–97

(Received 4/29/94, Revision accepted 12/20/94)

APPENDIX

In our analysis of motion coherence, we have employed a line element model (Watamaniuk, Sekuler, & Williams, 1989, Williams, Tweten, & Sekuler, 1991) that is built around the notion that the visual system represents orientation using a fixed number of tuned mechanisms Discrete representations of orientation are now commonly incorporated into models of motion perception (Wilson, Ferrera, & Yo, 1992), and psychophysical evidence suggests that there are about 24 such mechanisms (Phillips & Wilson, 1984) The visual computation of angle within a line element model is accomplished by N Gaussian mechanisms spaced $2\pi/N$ radians apart There is considerable freedom in the choice of the bandwidth (or standard deviation) of these mechanisms, and we follow Watamaniuk et al (1989) in choosing the half bandwidth at half maximum to be equal to the mechanism spacing Then the response of the ith mechanism to motion in a direction θ is

$$R_i(\theta) = e^{[-(\theta-\theta_i^c)^2/h^2 \ln 2]},$$

where θ_i^c is the angle at which the ith mechanism is centered, and $h = 2\pi/N$ is the half width at half maximum The response of a single mechanism to an entire stimulus representing a distribution of directions is

$$R_i = \int_0^{2\pi} R_i(\theta) p(\theta) d\theta,$$

where $p(\theta)$ is the probability that an element in a random-dot cinematogram (RDC) moves in a direction θ

The response functions, R_i contain all the information that the visual system has about the stimulus Formally, they are a discrete representation of the continuous distribution, $p(\theta)$, that is specified by the random stimulus The response function becomes a probability simply by normalizing with respect to the entire level of stimulation,

$$p_i = \frac{R_i}{\sum_{i=1}^{N} R_i}$$

The theoretical models presented in this article are based on various moments of the direction distributions used in the construction of RDCs The visual system does not have access to these distributions

per se, but only to the discrete representations that they engender Consequently, we have used discrete probabilities, p_i, in our theoretical models The discrete representations of the mean, variance, and information entropy are computed as

$$\mu = \sum_{i=1}^{N} p_i \cos\theta_i^c$$

$$\sigma^2 = \sum_{i=1}^{N} p_i \cos^2\theta_i^c - \mu^2$$

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i$$

These moments suffice for the mean, variance, risk, and entropy calculations that are presented here

It is of interest to inquire how the line element model would behave in a channel capacity experiment of the sort summarized by Miller (1956) or Hake and Garner (1951) We calculated the model's channel capacity by regarding $R_i(\theta_j)$ as being proportional to the probability that the model responds with the categorization θ_i^c upon being presented with motion in a direction θ_j The matrix elements $a_{ij} = R_i(\theta_j)$ form a confusion matrix that can be analyzed using the standard techniques summarized in Attneave (1959) The results of this calculation turned out to be simply summarized ($r^2 = 1$) by the linear relation

$$\text{channel capacity} = N/3\ 5,$$

where we are measuring capacity not in bits but as the number of distinct angle categories In the range $24 \leq N \leq 28$ that was required by the entropy model to fit the coherence data, the corresponding capacity is seven or eight categories This is consistent with Miller's general rule of 7 ± 2, and is in exact agreement with Pollack's measurement of the channel capacity for identifying angle of inclination at short exposures (unpublished results cited by Miller, 1956)