Title
Observing Behavioral Qualities

Journal
International Journal of Comparative Psychology, 10(4)

ISSN
0889-3675

Authors
Fagan, Robert
Conitz, Jan
Kunibe, Elizabeth

Publication Date
1997

DOI
10.46867/C4BP41

International Journal of Comparative Psychology, Vol. 10, No. 4, 1997

OBSERVING BEHAVIORAL QUALITIES

Robert Fagen, Jan Conitz and Elizabeth Kunibe
University of Alaska Fairbanks, USA

ABSTRACT: Animal movements have distinctive qualities, and these qualities can vary even when the form of the movement remains relatively constant. Description of behavioral qualities by trained observers can be useful in basic behavioral research and in applications ranging from behavioral ecology to clinical medicine. A method called Laban movement analysis differentiates four separate bipolar effort factors that contribute to the quality of body movement. Using independent rankings of videotaped behavioral sequences, we verified that observers can distinguish behavioral qualities reliably when using the Laban system. Observers generally agreed both on the kind(s) of effort factor(s) present and on the mode or degree of expression of each factor. We discuss the potential and limitations of the Laban system as applied to animal behavior and identify some philosophical issues that arise from attempts to link the study of behavioral quality to the study of form and space, and to a possible emerging "science of qualities".

Qualities of behavior vary in nature. A familiar individual's distinctive walk is recognizable at a distance, long before facial features or body markings become evident. A primate mother might reject her infant by pushing it away roughly or by pushing it away gently. Play behaviors in young and adult animals also appear to have distinct qualities. For example, play-gaits in rhesus macaques (*Macaca mulatta*) (Sade, 1973) and in other species (Fagen, 1981, p. 415) have a bouncy or gamboling quality. Observers of play report relaxed muscle tone (Aldis, 1975), lightness and lack of tenseness (Kaufman & Rosenblum, 1966), and exaggeration of individual movements (Loizos, 1966).

Qualities of an action must be appropriate to that action's goal. Precise painting with a brush and shaking paint off the same brush are actions having similar forms, different qualities, and different effects and functions. Hands may clap in anger or in delight -- the same gesture, expressive in different ways through contrasting qualities.

Address correspondence to Robert Fagen, Fisheries Division, University of Alaska Fairbanks, 11120 Glacier Highway, Juneau, Alaska 99801 USA.

Behavioral qualities are important for efficiency and expressiveness of functional behavior. Furthermore, an observer's pre-existing perceptual biases may make the qualities of a particular action attractive to that observer even though this action lacks direct biological function. For all these reasons, it is interesting to try to observe behavioral qualities (the "how" of behavior) reliably.

Pioneering observers of behavior noticed behavioral qualities and clearly sensed that these qualities were important for addressing the questions or phenomena under study. Ecologist Robert MacArthur (1958) informally contrasted the "nervous" foraging movements of the black-throated green warbler *Dendroica virens* with the "deliberate" feeding habits of the bay-breasted warbler *D. castanea*. Clinician and writer Oliver Sacks (1987) summarized classic signs of Parkinsonism including quick, abrupt, and brief movements ("festination" and "pulsion").

Formal methods for describing behavioral qualities are little-known, and scientists have not previously considered using such tools in their work. Laban movement analysis (Bartenieff, 1980; Dell, 1977; Laban & Lawrence, 1974; Moore & Yamamoto, 1988) is a long-established method for studying the form and qualities of movement. We hypothesized that it would furnish a reliable approach to qualities of nonhuman behaviors. Levy (1990) used Laban methods in an exploratory study of confined, trained Atlantic bottlenose dolphins *Tursiops truncatus* and their trainers. Zhou and Shirley (1997) included Laban terminology in their description of red king crab *Paralithodes camtschaticus* locomotion. Neither study measured interobserver reliability.

Laban and his successors analyzed movement in terms of Body, Effort, Space and Shape elements. Effort analysis specifically addresses dynamic qualities of body movement. The other three elements can contribute in various ways to expressive movement but do not address movement dynamics *per se*. The term "body" refers to initiation and sequencing of movements, and body parts involved in movement. "Space" defines different kinds of spatial pathways that an actively moving body can take. "Shape" defines ways in which the body forms itself in space.

METHOD

We used Laban Effort methods to observe animal behavior. Laban

Effort analysis identifies four motion factors for identifying specific qualities or dynamics of movement. These factors are Space, Time, Weight, and Flow. Each factor represents two contrasting ways or modes of expression in which muscular effort can produce movement.

Space can be direct or indirect. Direct space effort is channeled in a particular spatial direction or focused on one point in space. Indirect space effort is expansive and multifocal. For example, a cat may focus all its energy and attention in one direction as it intently watches a mouse hole (direct). A bear may stand bipedally, looking and sniffing around as it seeks to localize the source of an odor or sound (indirect).

Time can be quick or sustained. Quick time effort produces acceleration or condensing of time with movements that appear rushed or hurried. Sustained time effort delays and decelerates movement. A stalking cat's movements are sustained until the final, quick spring on the prey.

Weight can be strong or light. Strong weight effort appears when pressure is applied forcefully during movement. Light weight effort involves a decrease in applied pressure so that the movement is not forceful. A bear may use all of its weight when digging for plant roots (strong). Or it may handle a gravid female salmon very delicately to extract the roe, a preferred food (light).

Flow can be bound or free. Bound flow effort serves to control and restrict movement. Free flow produces easy, unrestrained movement (free). A barefoot child may walk carefully across a stony beach (bound) to run and splash with abandon in the water, arms and legs swinging in all directions as the inertia of the movement is allowed to take over (free).

In Laban's work, the four elements Body, Effort, Shape and Space are connected. However, each element may be examined individually (Andrews & Scott 1986). We chose to concentrate on Effort while recognizing that additional insights and accuracy regarding behavioral quality may result from use of additional elements of the full Laban system, particularly Shape, to analyze behavior.

Methods used here to study behavioral qualities address the "how" of behavior. Form and position of movement, by contrast, constitute the "what" of behavior. Precise methods for the study of form and position include Eshkol-Wachmann notation (Golani, 1992) and aspects of Labanotation (Hutchinson, 1970). Eshkol-Wachmann and Labanotation methods differ (Bartenieff, 1980; Horwitz, 1988). Neither method addresses behavioral quality.

Procedure

We investigated reliability of Laban analysis for behavioral quality, using two observers in a laboratory setting. Both observers were biology students in the senior author's introductory one-semester (2 h/wk for 16 wk) Laban analysis workshop. Neither student was a professional dancer or college dance major. The scoring session whose results are reported here took place at the end of the workshop.

Observers analyzed eighteen independent, non-contiguous segments of six videotapes (three segments per tape) showing behavior of nonhuman animals in naturalistic settings. No tape previously shown in the workshop was analyzed in these tests. Additional criteria for choosing test tapes were that (1) the technical quality of the video recording had to be adequate for analysis, (2) each tape had to be of a different nonhuman species, (3) no portion of any test tape had been analyzed previously in terms of movement qualities, (4) the segments analyzed should show animals moving on or from a solid substrate rather than in a fluid medium, and (5) extremely small and extremely large animals (relative to humans) were not included in this initial study because time, weight and space scales for movement in animals whose body size differs from that of humans by many orders of magnitude are relative and may require special study techniques, correction factors or methods of analysis.

We used the first six tapes that met all these criteria. Species and activities analyzed included one primate (*Gorilla gorilla* travel and social behavior), four carnivores (*Canis familiaris* bounding, object handling and vigilance, *Panthera pardus* pouncing and pawing, *Panthera tigris* travel, *Ursus arctos* jumping, prey handling play and locomotor play), two ungulates (*Capra ibex* social contact and travel, *Rupicapra rupicapra* pawing), and one bird (*Pica pica* displacement, travel and vertical movement). 1/2 inch Super Beta format tapes were played on a SONY SL-340 video machine and viewed on an 18-inch SONY KX-1901 monitor. Both observers watched and then scored each tape segment simultaneously. Observers were given as much time and as many repetitions as they needed to score each segment before going on to the next. Observers were not allowed to communicate with one another or to discuss their observations or conclusions until after all scores had been recorded for the entire session. Observers scored each segment independently of one another for the Laban motion factors Space (direct or indirect), Time (quick or sustained), Weight (strong or light), and Flow (bound or free). If one or more factors were absent then the segment was scored as "absent" for that factor or factors. For example, the first observer scored the first segment on the first videotape as indirect Space and quick Time, with Weight and Flow both absent.

Our primary measure of agreement between observers was Kendall's (1975, p. 45-48) measure of association in a contingency table with ordered categories. Kendall's measure is preferable to kappa coefficients (Landis & Koch, 1977) because Kendall's measure recognizes the ordering of modes of expression on each motion factor, whereas kappa does not. We also calculated kappa coefficients because they are more familiar to most students of behavior. The terms "fair", "moderate", "substantial" applied in this paper and elsewhere to kappa coefficients are arbitrary and refer to the intervals 0.21-0.40, 0.41-0.60, and 0.61-0.80 respectively (Landis & Koch, 1977, p. 165). No "magic figure" exists beyond which a behavioral measure is accepted as reliable (Martin & Bateson, 1986, p. 91). Important but difficult-to-measure categories with fair to moderate kappa coefficients can still be useful (e.g., Smith & Vollstedt, 1985). Correlations are not appropriate for measuring reliability in this particular study because data are categorical (nominal) rather than ordinal or interval.

Observers' responses could differ in kind and/or in degree. We say that responses differ in kind when both observers notice the presence of the same behavior quality factor but disagree about its mode of expression. For example, suppose that both observers noticed the presence of Flow, but one observer recorded bound Flow, the other free Flow in the same movement or movement sequence. In this case, the observers' responses would be said to differ in kind. But what if one observer noticed Flow (bound or free) but the other saw no Flow quality and scored Flow as absent in the same movement or movement sequence? In this case, we would say that the responses differed in degree. One observer considered Flow to be present and the other observer viewed it as absent. To analyze degree and kind separately, we calculated separate reliability coefficients for these two components of observer response (illustrative numerical example, Table 1).

We felt that two reliable observers would very seldom differ in kind. Both would usually agree that they saw the same mode of expression for Flow (for example), if intense Flow was truly present. One would not see bound Flow and the other see free Flow in the very same movement. We also realized that observers might sometimes disagree about presence or absence of a particular behavior quality. All factors of movement quality can occur to some degree in actual movements, and advanced Laban movement analysis distinguishes different levels or intensities of expression of a given factor (Dell, 1977, p. 12, 32). (See Appendix for further explanation).

Note that when agreement in kind is analyzed for eighteen video segments, the total number of agreements/disagreements for analysis will be less than or equal to eighteen because we are only considering those cases in which both observers scored a particular factor as present. For example, in Table 1 there were four cases in which at least one observer scored a particular factor as absent, leaving fourteen cases for analysis of agreement in kind.

Table 1. Illustrative Data from Two Observers, Analyzed in Terms of Agreement in Kind and Agreement in Degree. Abbreviations: I = indirect spatial focus; D = direct spatial focus; Y = present; N = absent.

		Observer 1								
		Original Data			Kind			Degree		
		I	-	D	I	-	D		Y	N
Observer 2	I	2	0	1	2	-	1	Y	13	1
	-	1	1	2	-	1	-	N	3	1
	D	1	1	9	1	-	9			

We did not test interobserver reliability coefficients for statistical significance. Level of statistical significance of a reliability coefficient says little about the actual degree of reliability (Martin & Bateson, 1986, p. 91).

RESULTS

Interobserver agreement on each factor was positive and fair to moderate for both measures (Table 2). Agreement in kind was positive, moderate to perfect, and stronger than agreement in degree (Table 3) for Time, Weight, and Space. For Flow, agreement in kind was moderate and agreement in degree was substantial.

Agreement in degree involves measures that allow data from different effort qualities to be combined. The number of observations in the combined sample is sufficiently large to allow G-squared tests of interobserver reliability (Fienberg, 1980). A hierarchical log-linear model in which the two observers' assessments of degree were associated, but independent of tape number, was the simplest model needed to fit the data (Table 4) adequately (G-squared = 20.4, $df = 15$, $p = 0.16$).

Table 2. Observer Ratings and Reliability. Abbreviations: I, indirect; D, direct; S, sustained; Q, quick; L, light; H, strong; B, bound; F, free.

	Observer 1															
	Space				Time				Weight				Flow			
		I	-	D		S	-	Q		L	-	H		B	-	F
Observer 2	I	2	0	1	S	2	2	0	L	3	4	0	B	4	1	2
	-	1	1	2	-	3	3	1	-	1	6	1	-	0	2	0
	D	1	1	9	Q	2	1	4	H	0	1	2	F	2	0	7
Kendall			.45				.34				.46				.44	
Kappa			.38				.27				.37				.54	

Table 3. Kappa Coefficients for Agreement in Kind and in Degree

Source of Agreement	Factor			
	Space	Time	Weight	Flow
Kind	0.56	0.50	1.00	0.44
Degree	0.15	0.24	0.22	0.77

Table 4. Three-Way Contingency Table for Reliability Analysis (Degree Only). Variable 1, rating of observer 1; variable 2, rating of observer 2; variable 3, tape number.

		Observer 1											
		Tape 1		Tape 2		Tape 3		Tape 4		Tape 5		Tape 6	
		Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
Observer 2	Y	7	0	7	1	6	2	8	4	7	1	6	2
	N	1	4	1	3	3	1	0	0	1	3	3	1

The non-significant P in this test means that the model fits the data (lack of fit not significant). A second model assuming no association among the three variables did not fit the data (G-squared = 29.8, df = 16, p = 0.02). Fits of the first and second models differed significantly (G-squared = 9.4, df = 1, $p < 0.005$). Interaction terms involving tape number did not improve the fit significantly when added to the first model.

Separate analyses of agreement in kind for each effort quality were not possible for reasons of sample size. However, Laban theory identifies indirect, sustained, light and free as "indulging" qualities and direct, quick, strong and bound as "fighting" or "resisting" qualities (Bartenieff, 1980, p. 51; Moore & Yamamoto, 1988, p. 157). We used this identification to construct a combined sample (Table 5), including all data and measuring both agreement in kind and agreement in degree.

Table 5. Three-Way Contingency Table for Reliability Analysis (overall). Abbreviations: I, indulging; F, fighting. Variable 1, rating of observer 1; variable 2, rating of observer 2; variable 3, tape number.

		Observer 1																	
		Tape 1			Tape 2			Tape 3			Tape 4			Tape 5			Tape 6		
		I	-	F	I	-	F	I	-	F	I	-	F	I	-	F	I	-	F
Observer 2	I	4	0	0	3	1	1	2	2	1	1	2	1	4	1	1	1	0	0
	-	0	4	1	1	3	0	2	1	1	0	0	0	1	3	0	1	1	2
	F	0	0	3	0	0	3	0	0	3	1	2	5	1	0	1	0	2	5

A hierarchical log-linear model in which the two observers' assessments were associated, but independent of tape number, was the simplest model needed to fit the data adequately (G-squared = 48.1, $df = 40$, $p = 0.18$). A second model assuming no association among the three variables did not fit the data (G-squared = 81.4, $df = 44$, $p < 0.001$). Fits of the first and second models differed significantly (G-squared = 33.3, $df = 1$, $p < 0.0001$). Interaction terms involving tape number did not improve fit significantly when added to the first model. Expectations were smaller than ideally required (48 of 54 were 0.67 or greater, 24 of 54 were 1.00 or greater). Expectations for the analysis of degree in kind would have been even smaller, so this analysis was not done.

DISCUSSION

Mastery of Laban analysis can require several years of full-time study. But the Laban techniques investigated here appear reliable when applied to animal behavior even by relatively inexperienced observers. When observers agreed that a particular factor was present, they generally agreed on its mode of expression.

To model the stream of ongoing behavior with a sequence of discrete patterns that are fixed in form and that vary only in frequency, duration and intensity (Fagen & Young, 1978) is to ignore qualitative aspects that a perceptive observer might well notice and consider important. The study of behavior can include both direct, objective data on acts in sequence and reliable descriptions of qualities that we notice in everyday life (cf. Feaver, Mendl, & Bateson, 1986; Hinde, 1976).

Subtle differences in movement qualities can determine how well one species does relative to another in a particular habitat. Predator avoidance is one component of such success, and ecological studies of vigilance behavior might productively distinguish between direct and indirect spatial effort by a vigilant animal. Behavioral qualities could play a role in ritualization, in perceptual movement cues, and in dynamic traits of visual communication, e.g. in studies of sexual selection and prey-predator interaction. Qualitative measures of drug effects in behavioral pharmacology could use Laban techniques as a novel exploratory tool. Behavioral qualities defined by Laban methods could help define individual distinctiveness. Laban methods may also prove useful in defining and measuring qualities of animal play.

Highly-skilled Laban observers can readily distinguish varying degrees of intensity along the axis of each effort quality. For example, one observer may score a movement as weakly Light, whereas another might score the same movement as weakly Strong. This pair of scores would have been analyzed as a failure to agree in degree, but would actually reflect a higher level of concordance than would occur from a complete disagreement in degree (very Light vs. very Strong). Because this study involved trained but not expert observers, we chose not to differentiate the scale of observation so finely. Had we been prepared to score degrees of intensity along each effort gradient, the resulting agreement in degree might actually have been higher than that reported here.

We were all trained by the same Laban teachers, but different schools of Laban studies exist and observers who have learned Laban techniques differently may not agree as well as we did. Our impression is that the differences between these schools involve matters of emphasis (of Effort versus Shape, for example) or simply of terminology.

The Laban approach has other obvious limits as well. Additional dimensions or sets of dimensions may prove fruitful for defining qualities of animal behavior. Problems of interpretation may confront analyses of very small and very large animals, and of animal movement in fluid media.

In Laban studies, observations of Effort are closely linked with observations of Shape, "how the body forms itself in space" (Dell, 1977, p. 43). Effort qualities are said to be easier to observe accurately when the observer also records Shape dynamics. This aspect of Laban methods merits consideration for future studies of behavioral quality.

Laban methods require a preliminary training period involving active movement by the prospective observers and motor learning of each relevant quality and concept. In general, can training programs that include active movement in a structured, systematic context actually help humans learn to observe animal behavior more reliably? This hypothesis deserves empirical investigation.

Laban methods have been applied to the study of dance and human movement across many cultures, but these methods may include language-specific or culture-specific features. Although Laban's concepts have successfully survived translation from German to English, it would be worthwhile to see just what native speakers of other languages, particularly non-Western languages, might offer regarding concepts and terms for movement quality. Valuable both in its own right and as an exercise in cross-cultural extension of the approach would be a study that compiles lists of qualititative traits and then determines reliability of these traits in a controlled manner (e.g., Fagen & Fagen, 1996; Feaver, Mendl, & Bateson, 1986). One or more individuals familiar with the animals selects a list of items that reflect a broad range of the types of behavioral qualities that occur under the conditions of observation. These items are then ranked by independent observers and the rankings compared statistically using nonparametric correlation methods.

The philosophical concepts of form, quality, and space play key roles in scientific theory and practice, as well as in the philosophy and making of art. Western concert dance, for example, admits a clear distinction between the form of a dance step from the classical syllabus and the various expressive qualities with which it can be performed. Analogous issues arise in music performance. A musical score can indicate both the notes to be played and the expressive quality intended by the composer. Expansive biological discussions of form, quality and space can also furnish grist for the philosopher's mill, as the following examples may indicate.

Goodwin (1996, p. 196-8) predicts that "a science of qualities" will emerge from the study of biological form (both spatial form, e.g. the shape of a salamander body, and temporal form, e.g. "the distinctive flight of a woodpecker"). Woodpeckers' undulating flight is indeed distinctive (Peterson 1980, p. 188). "Distinctive" here seems to imply the form of the spatial path traced by the bird's movements, but flying or any sort of locomotion will generally involve effort qualities as well. Laban sought to relate dynamics or qualities of movement to qualities of form as expressed in shape, in space, and in the moving body's dynamics of muscular exertion or effort.

Concepts of design (e.g., Smith 1981, Stevens 1974) endure in discussions of biological form and natural pattern. Haeckel coined the term "ecology" from the Greek *oikos*, "house". Grinnell named and defined the ecological niche. Architectural spandrels (Gould & Lewontin, 1979; Gould, 1997) inspire evolutionary ideas. Schwenk (1976) viewed form as the visible manifestation of dynamic processes, and especially of hydrodynamics. Organisms "express their natures through the particular qualities of their form in space and time" (Goodwin 1996, p. 198). Is it meaningful to suppose that ways of thinking intrinsic to a science of qualities might resemble ways that designers of all persuasions, from architects to choreographers, work and play with qualities of form in space and time?

ACKNOWLEDGEMENTS

We thank Betsy Kagan and Liz Shipman for introducing us to Laban Movement Analysis, Barbara Levy and Jackie Hand-Vagorio for information on Laban studies of dolphins, Brynar Mehl, Matthew Nash, Colleen Dilts, Angela Fowler, and Nancy Ratner for discussions and demonstrations of movement qualities in dance, and Betsy Kagan, Peter Klopfer, and Klopfer's graduate lab discussion group at Duke University as well as two anonymous referees for comments on drafts of the manuscript.

REFERENCES

Aldis, O. (1975). *Play fighting*. New York: Academic Press.

Andrews, M., with Scott, C.B. (1986). The Bartenieff fundamentals (TM). *Contact Quarterly*, Spring/Summer.

Bartenieff, I. (1980). *Body movement*. New York: Gordon & Breach.

Dell, C. (1977). *A primer for movement description* (2nd ed.). New York: Dance Notation Bureau Press.

Fagen, R. (1981). *Animal play behavior*. New York: Oxford University Press.

Fagen, R., & Fagen, J.M. (1996). Individual distinctiveness in brown bears, *Ursus arctos L. Ethology, 102*, 212-226.

Fagen, R., & Young, D.Y. (1978). Temporal patterns of behaviors: Durations, intervals, latencies and sequences. In P. Colgan (Ed.) *Quantitative ethology* (pp. 79-114). New York: Wiley-Interscience.

Feaver, J., Mendl, M., & Bateson, P. (1986). A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour, 34,* 1016-1025.

Fienberg, S.E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, Mass.: MIT Press.

Golani, I. (1992). A mobility gradient in the organization of vertebrate movement: The perception of movement through symbolic language. *Behavioral and Brain Sciences, 15,* 249-308.

Goodwin, B. (1996). *How the leopard changed its spots.* New York: Touchstone.

Gould, S.J. (1997). The exaptive excellence of spandrels as a term and prototype. *Proceedings of the National Academy of Sciences USA, 94,* 10750-10755.

Gould, S.J., & Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society, London, B, 205,* 581-598.

Hinde, R.A. (1976). On describing relationships. *Journal of Child Psychology and Psychiatry, 17,* 1-19.

Horwitz, D.L. (1988). Philosophical issues related to notation and reconstruction. *Choreography and Dance, 1,* 37-54.

Hutchinson, A. (1970). *Labanotation.* New York: Theatre Arts Books.

Kaufman, I.C., & Rosenblum, L.A. (1966). A behavioral taxonomy for *Macaca nemestrina* and *Macaca radiata*, based on longitudinal observation of family groups in the laboratory. *Primates, 7,* 205-258.

Kendall, M.G. (1975). *Rank correlation methods* (4th ed.). London: Charles Griffin.

Laban, R., & Lawrence, F.C. (1974). *Effort* (2nd ed.). Boston: Plays, Inc.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Levy, B.A. (1990). *Psychoaesthetics and dolphin personality.* Unpublished master's thesis. Pratt Institute, Brooklyn, NY.

Loizos, C. (1966). Play in mammals. *Symposia of the Zoological Society of London, 18,* 1-9.

MacArthur, R.H. (1958). Population ecology of some warblers of northern coniferous forests. *Ecology, 39,* 599-619.

Martin, P., & Bateson, P. (1986). *Measuring behaviour.* Cambridge: Cambridge University Press.

Moore, C.-L., & Yamamoto, K. (1988). *Beyond words.* New York: Gordon & Breach.

Peterson, R.T. (1980). *A field guide to the birds.* 4th ed. Boston: Houghton Mifflin.

Sacks, O. (1987). *Awakenings.* New York: Summit Books.

Sade, D. S. (1973). An ethogram for rhesus monkeys. I. Antithetical constrasts in posture and movement. *American Journal of Physical Anthropology, 38,* 537-542.

Schwenk, T. (1976). *Sensitive chaos.* NY: Schocken Books.

Smith, C.S. (1981). Structural hierarchy in science, art and history. In J. Wechsler (Ed.) *On aesthetics in science* (pp. 9-54). Cambridge, Mass.: MIT Press.

Smith, P.K., & Vollstedt, R. (1985). On defining play: an empirical study of the relationship between play and various play criteria. *Child Development, 56,* 1042-1050.

Stevens, P. (1974). *Patterns in nature.* Boston: Little, Brown.

Zhou, S., & Shirley, T.C. (1997). Behavioural responses of red king crab to crab pots. *Fisheries Research, 30,* 177-189.

APPENDIX

Measuring Separate Components of Agreement with Kappa

1. Agreement in kind. To calculate a kappa coefficient that measured observers' agreement on the mode of expression of Flow (for example), given that both observers agreed on the presence or absence of Flow in the movement, we calculated expected values for a 3x3 table that had structural zeroes in the four cells (1,2), (2,1), (2,3), (3,2) where exactly one observer scored the Flow as absent. This arrangement of structural zeroes ensured that expected values only contributed to kappa when observers agreed that Flow was present, or when they agreed that Flow was absent. We made these calculations with the loglin() function in S-Plus for Windows, version 3.3 (StatSci Division, MathSoft Inc., Seattle, Washington, USA), which implements iterative proportional fitting methods for incomplete contingency tables (Fienberg, 1980).

2. Agreement in degree. To calculate a kappa coefficient that measured observers' agreement on the presence or absence of a particular motion factor, we calculated expected values for a 2x2 table constructed as follows. For the upper left-hand corner of the table, cell (1,1), we counted all cases in which observers agreed that Flow (for example) was present. For the lower right-hand corner of the table, cell (2,2), we counted all the cases in which observers agreed that Flow was absent. Off-diagonal cells contained the counts of cases in which one observer saw Flow and the other did not.