Problems of behavior measurements

Ilan Golani[1] and Yoav Benjamini[2]
[1]*Departments of zoology, Tel Aviv University*
[2]*Department of Statistics and Operation Research, Tel Aviv University*

The suggestion that a "good" description of behavior is what a good ethologist considers to be a good description highlights a necessary but not sufficient condition for high-quality descriptions. This is because also good ethologists can err. How, then, if not on the basis of a claim for authority, can one tell the difference between high and low quality descriptions? One help comes from behavior genetics, where differences in behavior are used for the localization of genes on chromosomes. The QTL method used for such analysis depends critically on the quality of the measurements of behavior. If under the same standard conditions one inbred mouse strain is found in one laboratory to be significantly higher on a behavioral measure than another strain, and found to be significantly lower on the same measure in another laboratory, then the results are *not* replicable across laboratories and therefore useless for gene localization studies. Statistical replicability across laboratories thus becomes an objective yardstick for both the relevance of a behavioral measure and for the estimation of the quality of its measurement[1,6].

The high benchmark required for obtaining replicability, demands high quality data. Obtaining such data is, therefore, not a luxury but a constraint dictated by the requirement of replicability. In kinematic studies this implies extensive preparation of the data for analysis, including elaborate smoothing[5] and data segmentation procedures[2] (http://www.tau.ac.il/~ilan99/see/help/), since using first, second, or even third derivative measures like velocity, acceleration, jerk, and curvature drastically increases the system's noise, thus putting severe demands on the quality of these procedures.

The raw data of the movement material collected by tracking systems are kinematic variables such as the time series of location data and their respective calculated derivatives (at the path scale), and movements of the parts of the body (interlimb coordination at the joints scale).

Analysis reveals that these variables sometimes form discrete patterns. Discrete patterns thus constitute the results of the study, not its beginning. If these patterns have a fixed content, then this content can be described only once, for all patterns, and analysis can proceed by using these identical patterns as the building blocks of behavior. However, in the majority of cases the content of these patterns is variable. A premature encapsulation of kinematic features into such patterns, whether by a human observer or by a neural network trained by a human observer, yields "behavior patterns" whose variable content becomes inaccessible for further analysis. These packaged and labeled building blocks may be useful for counting frequencies in time and space, but they constitute "black boxes" as far as a moment-to-moment dynamic analysis is concerned. Hence the numerous ethograms - lists of inert labeled behavior patterns left by classical ethology, which are useful as first-approximation-descriptions but useless as far as the comparative dynamics of behavior are concerned.

In the current computational age, segmentation and packaging of the stream of behavior into discrete patterns is, fortunately reversible and therefore not problematic. Since the time-series of kinematic data are indexed, segmentation is performed at the indexing level, leaving the kinematic time-series intact and accessible for any other type of analysis or any other type of segmentation. Dissecting the flow at the indexing level preserves the transparency of the patterns and allows one to segment the flow in several compatible ways, each highlighting other aspects of the organization of behavior. For example, in our studies of mouse exploratory behavior we segment the path traced by the mouse into lingering episodes and progression segments, based on their speed profile[2]. On the one hand, the topographical distribution of lingering is used to define preferred places[4], and the probability of their performance at particular locations is used to define locational memory[3]. On the other hand, the speed profile within lingering episodes is used to calculate average lingering speed[6] – a highly heritable and discriminative measure characterizing the level of activity during staying-in-place behavior across strains and preparations. By examining the content of lingering episodes at the joints scale, one can also readily establish the momentary level of familiarity the mouse has with the novel environment: in a novel environment mice perform horizontal head scans, whereas in a familiar one they tend to also perform vertical scans. Furthermore, a scan in a particular direction often forecasts the direction the mouse is going to take next. The mouse's location is disclosed at the path scale, whereas the direction of its attention and its intentions (see Fentress, this symposium) are disclosed at the joints scale. Segmentation thus delineates the high-level units, whereas the dynamic content of these units relates them to history or topography, or modulates their significance. The higher level progression segments and lingering episodes can further be assembled into incursions (forays into the center) and excursions (roundtrips from home base), which in turn can be used to define higher level constructs like familiarity[8], locational memory[3], and anxiety[7].

References

1. J. C. Crabbe, D. Wahlsten, B. C. Dudek (1999), *Science* 284-1670.
2. Drai D, Benjamini Y, Golani I. Journal of Neuroscience Methods. 96 (2) (2000)
3. Dvorkin A, Benjamini Y, Golani I (2008) *PLoS Comput Biol* **4(2)**.
4. Golani I, Benjamini Y, Dvorkin A, Lipkind D, Kafkafi N. In: The Behavior of the Laboratory Rat – A Handbook with Tests Edited by Ian Q. Whishaw and Bryan Kolb. Oxford University Press.
5. Hen I, Sakov A, Kafkafi N, Golani I, Benjamini Y. (2004) *Journal of Neuroscience methods* **133** (1-2): 161-172.
6. Kafkafi N, Benjamini Y, Sakov A, Elmer GI, Golani I. (2005) *Proc Natl Acad Sci U S A*. 22;102(12):4619-24..
7. Lipkind D, Sakov A, Kafkafi N, Elmer GI, Benjamini Y, Golani I. (2004). *J Appl Physiol.* **97(1)**:347-59.
8. Tchernichovski, O., Benjamini, Y. (1998) *Biological Cybernetics* **78(6)** 433-440