The Advantages of Phenotyping Mouse locomotor Behavior Simultaneously in Multiple Laboratories

Coworkers: Neri Kafkafi, Roy Wise (NIDA, Baltimore),Greg Elmer (MPRC, Baltimore), Yoav Benjamini, Dina Lipkind, Guy Horev, Michal Pagis, Ehud Fonio, Anna Dvorkin (Tel Aviv), Itay Hen, Nir Lev (Programmers, Tel Aviv)

Presented by Ilan Golani,Tel Aviv University, at the *Half Moon Bay "New Mouse Behavioral Phenotyping Assays and Mutagenesis for Brain and Behavior Phenotypes December 11-13, 2001*.

Introduction

Two current key issues in behavioral phenotyping are discrimination across strains and replicability across laboratories. The Crabbe Wahlsten & Dudek paper appearing in *Science (1999)* showed that in spite of a detailed experimental protocol equated across laboratories, phenotyping results are often idiosyncratic to particular laboratories. This conclusion presents a challenge not only to the field of behavior genetics but also to the field of ethology! How is it that locomotor behavior, for example, a *par excellence* innate pattern is not replicable across laboratories?

Our response to this challenge has both ethological (part I, presented by Golani) and statistical (part II, presented by Benjamini) aspects.

SEE: To obtain both discriminative and replicable results we use SEE (a software supported Strategy for the Exploration of Exploration, Drai & Golani, 2001). Main features of this strategy are:

1. Using a large (2.5m diameter) circular arena. Fast strains should clearly have the opportunity to manifest their speed and acceleration potentials, thus providing a better differentiation from their slower counterparts.
2. Using long (1/2h) sessions. Some strains (e.g., Balb) show a build up in activity whereas others (e.g., C57) show a shut down (Drai et al., submitted). The mice need time to manifest this and other differences in across-the-session-dynamics.
3. Using an appropriate resolution of the phenomenon (25-30 Hertz). As in histology, structure should be examined under appropriate microscope magnification. A low sampling rate implies losing stops whose duration is as short as 0.2s. Because strains differ substantially in stopping behavior (and in the form and duration of scanning during stopping), discrimination and replicability are improved by capturing these stops.
4. Like anatomists we carefully cleave the complex structure into naturally distinct "tissues". Locomotor behavior consists of an agglomeration of patterns. Therefore we first separate the behavior into distinct classes of patterns.
5. Only then do we compute the endpoints for each class of patterns separately (as in neurochemistry, where neurotransmitter levels are established in distinct tissues rather than in a homogenate of the whole brain).

Cleaving the structure into distinct tissues*:* Consider the example presented in figure 1, of a path traced by a DBA mouse in the course of half a minute*:*

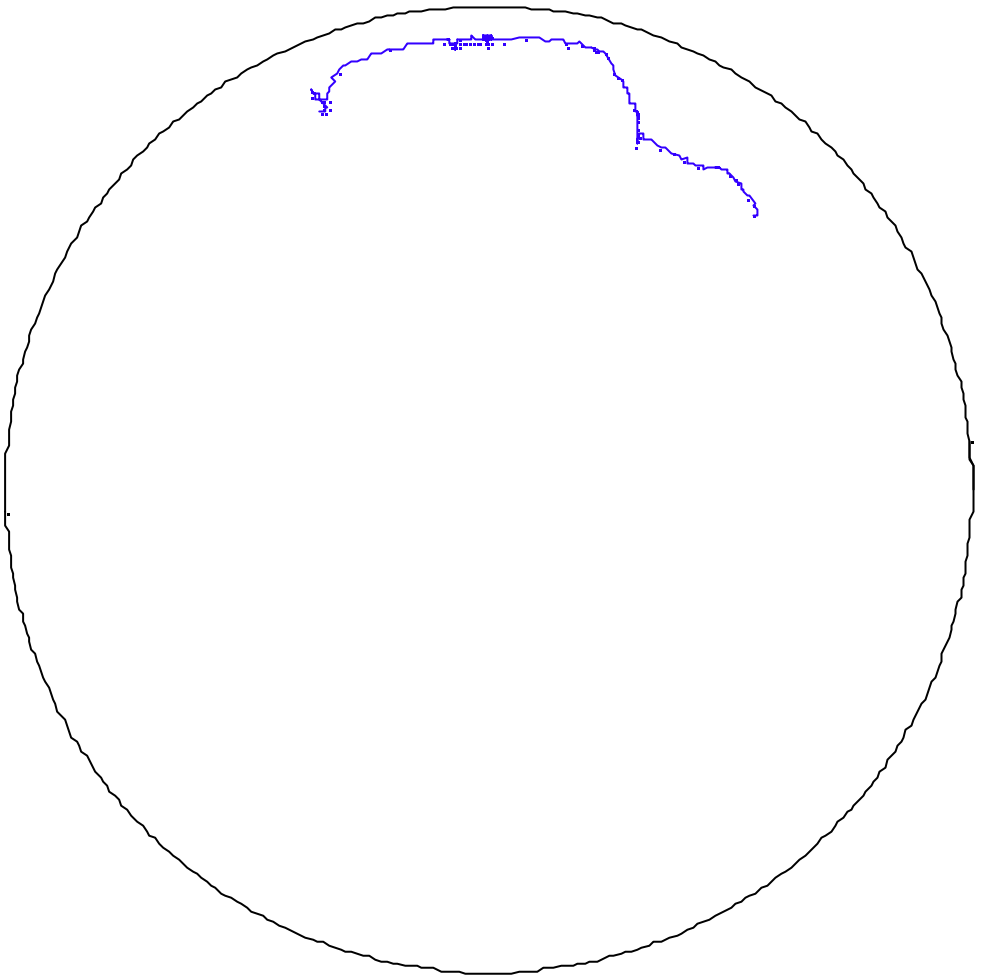

Figure 1.

All you can get from the raw path traced by a mouse in the arena (figure 1) is the distance traveled, the proportion of this distance traveled along the wall and away from it, and the amount of winding of the path. Note that the trace is also quite noisy. To compute the speed of the animal along this path we calculate the first derivative of the time series of the location data, and because each value in the computed speed time series is obtained from two (noisy) location data points, the graph is several folds more noisy (figure 2):

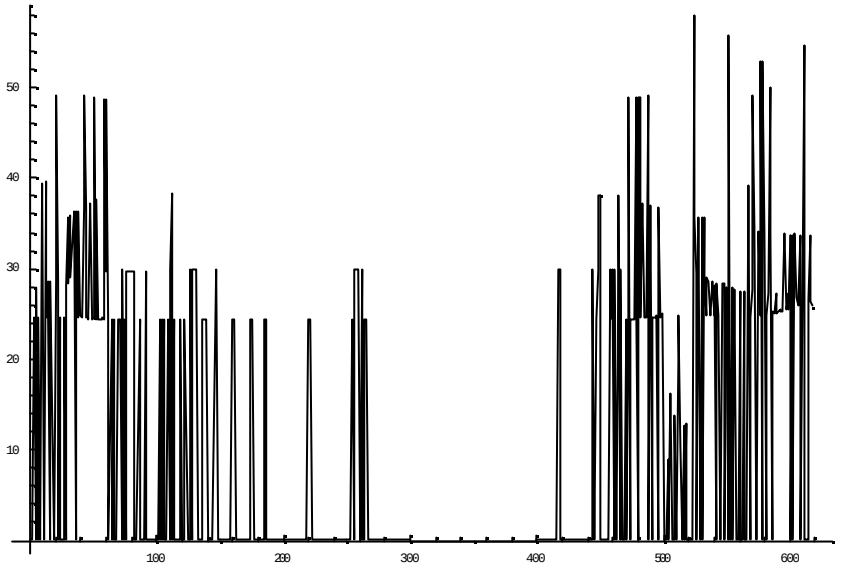

Figure 2.

We thus must smooth the path, and we do so by using SEE Path Smoother (SPSM; Benjamini et al., 2001), a stand alone, user friendly publicly available software developed by us for this purpose. This software combines the use of a running median smoother that captures even the shortest stops, and a subsequent smoothing method called Lowess which assures a robust smoothing (Hen et al., in preparation; see figure 3)

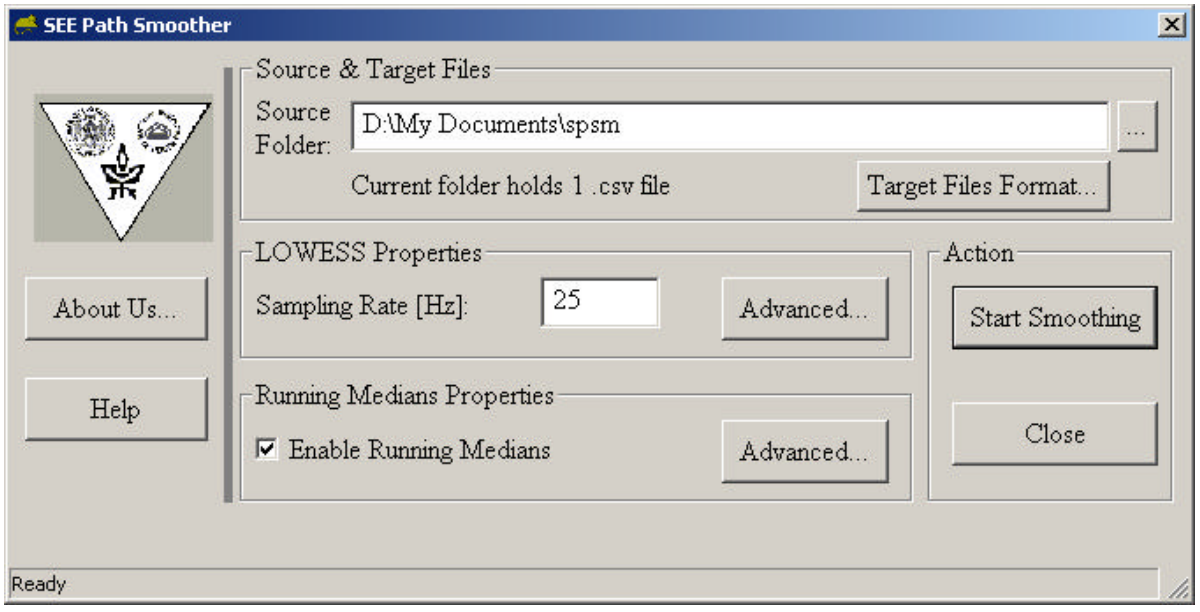

Figure 3.

After smoothing, the speed time series presented in figure 2 takes the following form (figure 4):

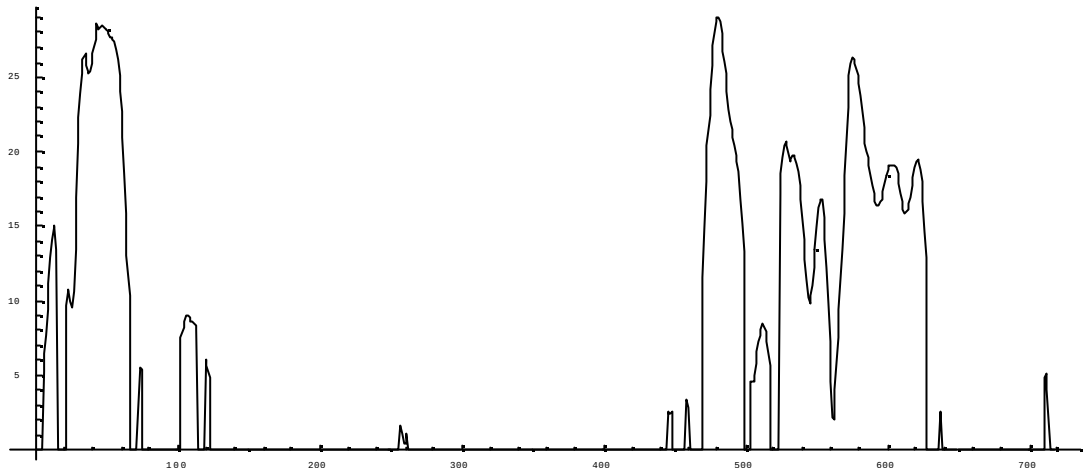

figure 4:

On a background of arrests (segments of zero speed), one gets the impression that there are, in the time series presented above, two distinct classes of motion segments – those with high speed, and those with low speed.

To examine this impression we obtain, for each movement episode, the peak speed reached in it, and plot a density graph (a sliding window histogram) of the frequency distribution of peak speeds in these movement episodes. As illustrated in the density graph of our mouse-session (figure 5), there are indeed two distinct populations: low speed segments (which we call lingering or stopping episodes) and high speed segments (which we call progression segments).

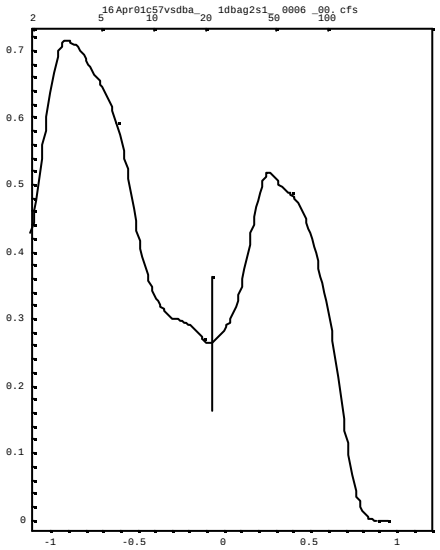

Figure 5

The segmentation is accomplished with SEE Path Segmentor (SPSG) (for rationale beyond this software see Drai et al., 2000). This stand alone, user-friendly, publicly available software was also developed by us for obtaining the threshold that distinguishes between the two types of segments (figure 6). The input for this software is the output of SPSM.

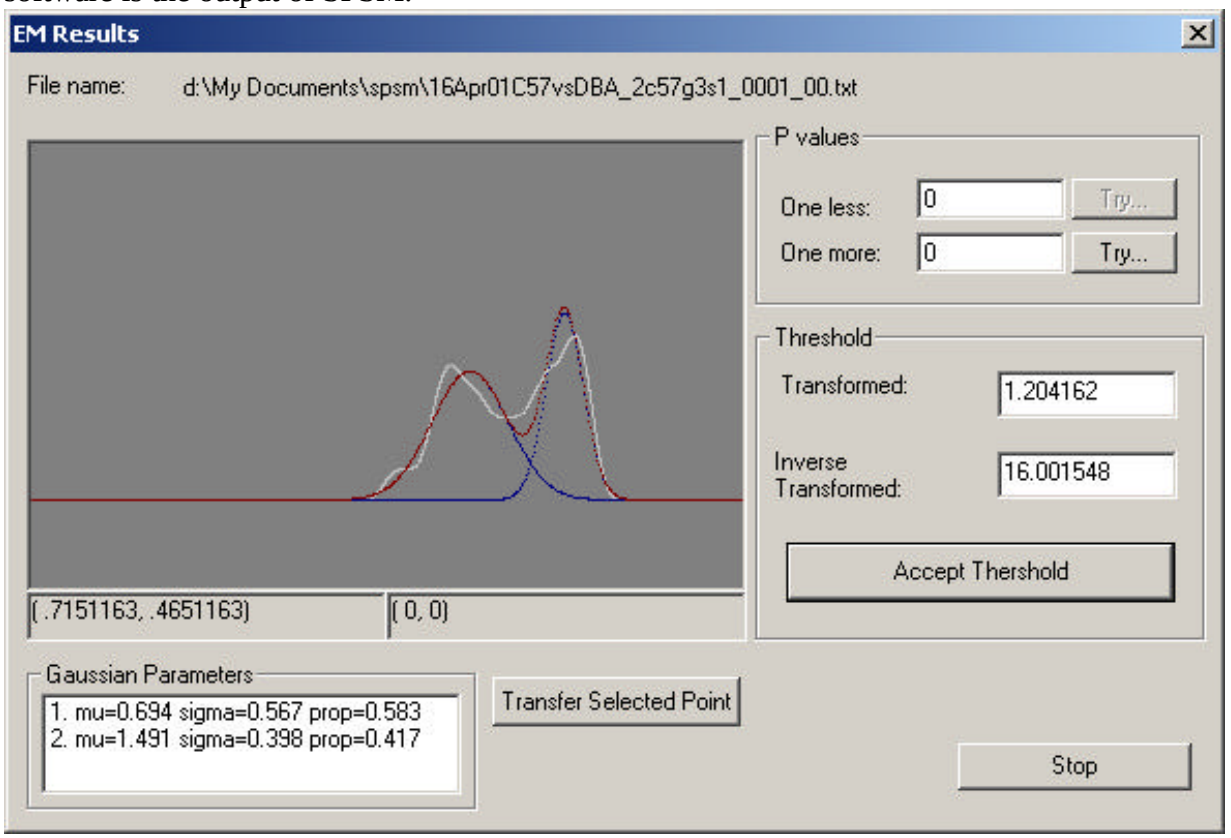

Figure 6:

A Video clip of a mouse performing the 1/2min behavior slab presented in figures 1,2, and 4 can be viewed at http//www.tau.ac.il/~ilan99. It reveals the richness of patterns underlying the path: the mouse lingers, performs a progression segment in a clockwise direction, stops, crawls slowly clockwise for two steps, grooms, performs a vertical scan, another progression segment, a short lingering episode, and finally a progression segment terminating with a lingering episode away from the wall. All these patterns are indirectly represented in the spatial spreads and durations of the segments automatically computed and traced in figure 7. In particular, note the long spatial spread of the second lingering episode (from left) which included stepping, grooming and scanning.

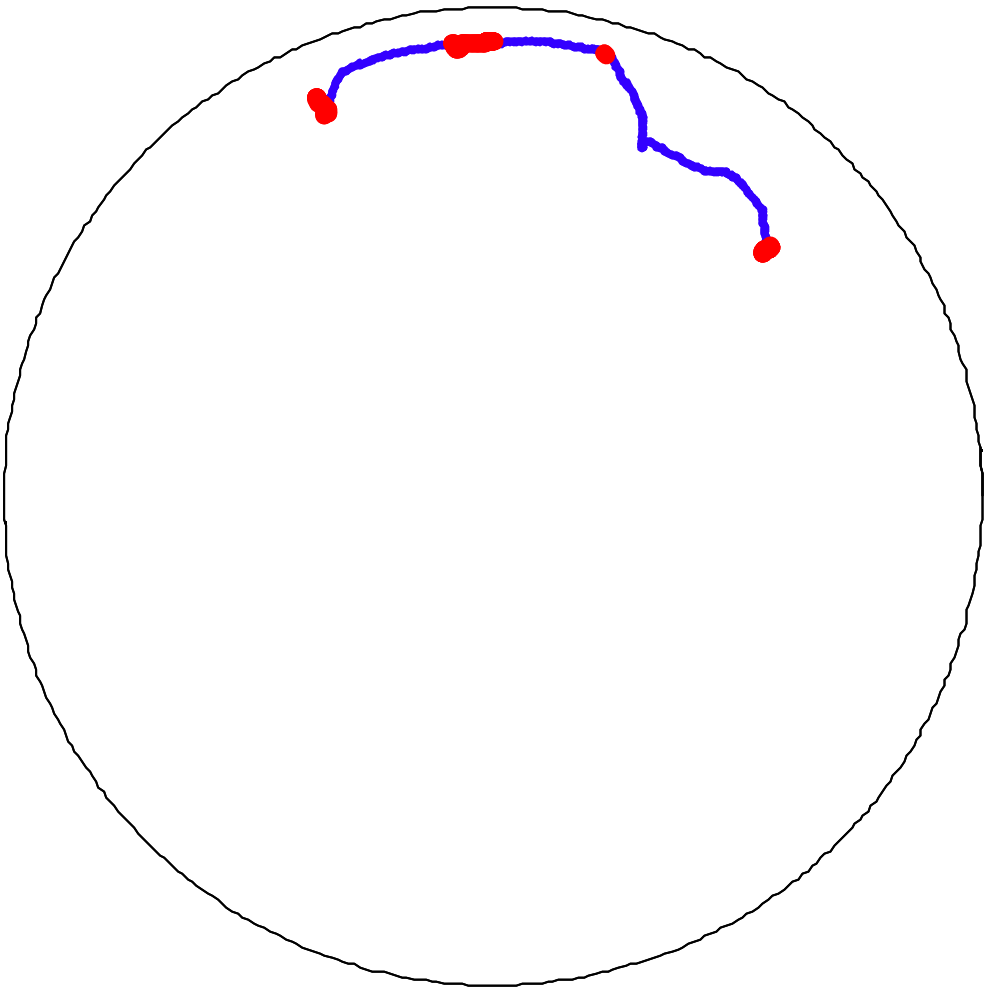

Figure 7.

The same segmented path with its corresponding speed trajectory (figure 8) changes the sequence into an information-rich sequence on which, not unlike a protein or DNA sequence, a large number of measurements can be performed. This sequence is, furthermore, stored on the web for future measurements of newly defined endpoints and for future comparisons with newly generated preparations and strains.

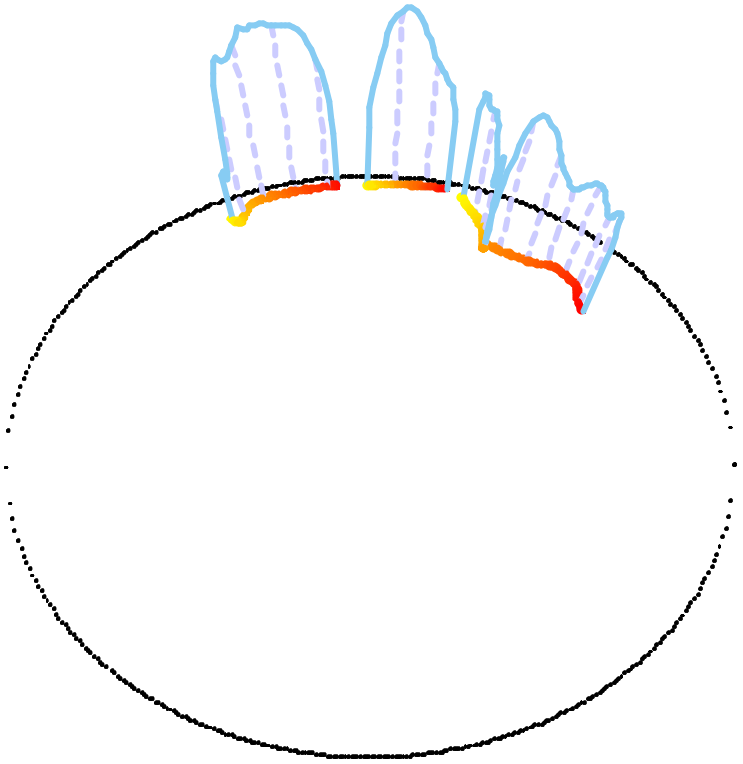

Figure 8: The yellow- to- red lines represent progression segments and the respective color transition designates the direction in which they are performed. Azure lines designate speed trajectories, and empty spaces between segments designate the spatial spread of lingering episodes.
Having obtained this articulated representation, all kinds of measurements can be performed. In particular, we can now compute the various endpoints for lingering (figure 9, right, red) and progression segments (figure 9, left, blue) separately, increasing thereby the likelihood of yielding more discriminative and replicable results.

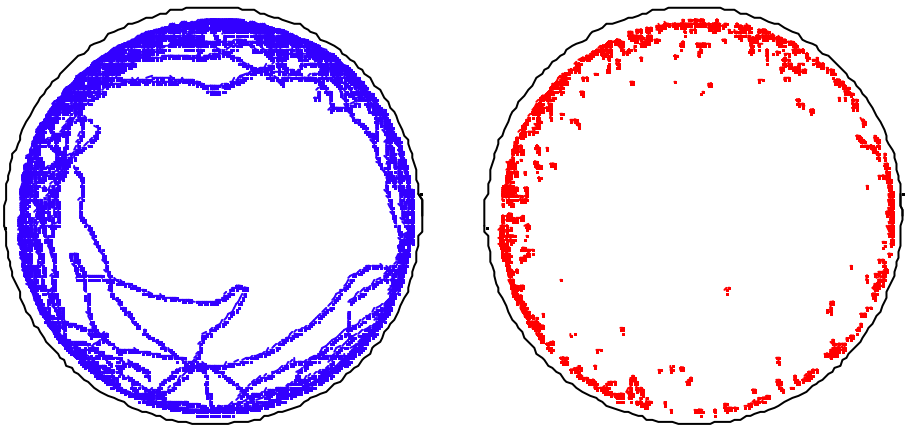

Figure 9

The advantages of working simultaneously in multiple laboratories

Rather than designing a behavioral endpoint in one laboratory and only then proceeding to challenge its replicability and power of discrimination across other laboratories, the multiple laboratories setup provides an opportunity to build this setup into the design of endpoints right from the start.

First example: Computation of maximal speed in near -wall progression segments (Dina Lipkind, Tel Aviv University)

In our first example, taken from the work of Dina Lipkind of Tel Aviv University (Lipkind et al., in preparation), the goal was to compare between strains the Maximal Speed obtained within progression segments. As with the design of other endpoints, we first want to make sure that the comparison will be performed on the right classes of segments, not mixing, as it were, “apples and oranges”.

The PathPlots of a DBA mouse (henceforth left in all figures) and a C57 mouse (right) represent the data before their partitioning into distinct classes of patterns (figure 10):

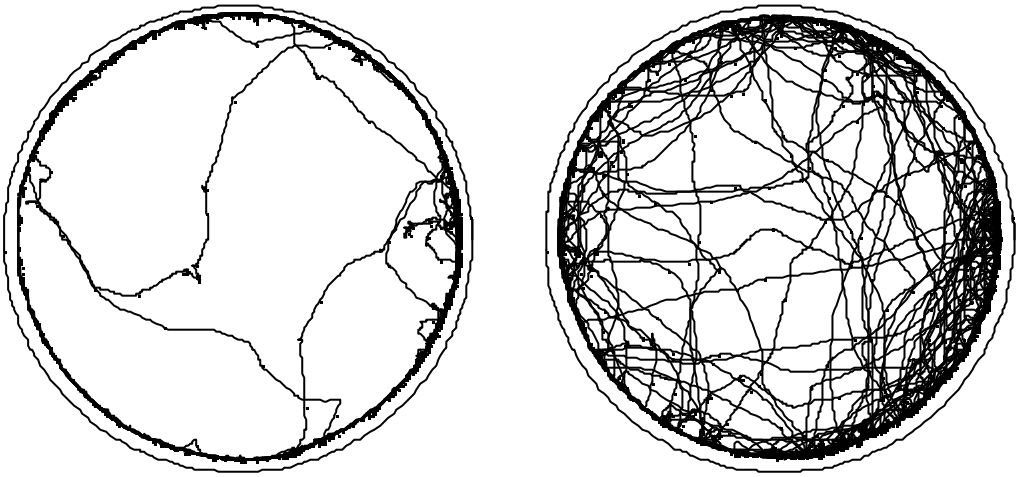

Figure 10

Figure 11 presents the Tel Aviv results of this comparison for 10 mice in each group:

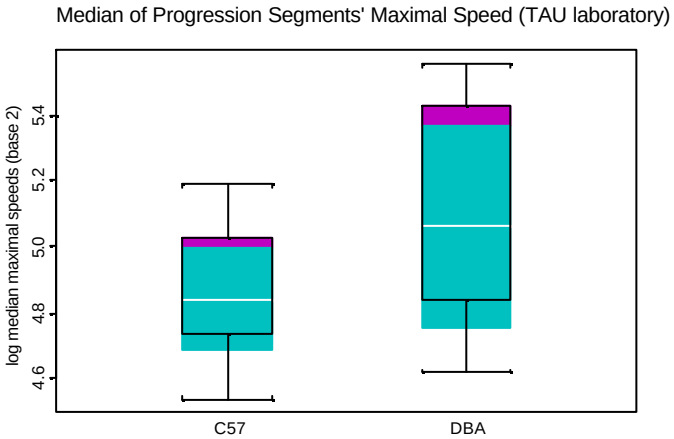

Figure 11: The white bars in the boxplot summaries of the results designate the group medians, the box below this bar represents the second quartile, and the box above, the third quartile. The vibrisae represent the group's maximum and minimum. Overlap of the turquoise colored areas (range of standard error) of the two boxplots implies absence of significant differences.

As shown, the difference between the strains was not significant. Having the impression that there is a substantial difference in the speed of progression of the two strains along the wall, but not necessarily in the center, Lipkind proceeded to classify progression segments according to their radial speed component (progression along wall being characterized by near-zero radial speeds).
Density graphs of log radial speed during progression indeed cleave into two sub-classes of segments: those with radial speed ~ 0 (Gaussian on left side of deep in each of the density graphs), and those with radial speed ~ > 0 (Gaussians on right of deeps).

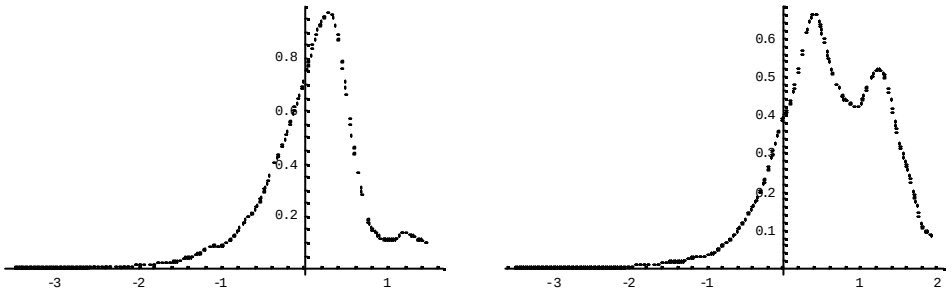

Figure 12. Log of radial speed on the x-axis enhances the deep between the two Gaussians.

Using the minima of the deeps as the thresholds for distinguishing between the two classes (and subsequently adding to the right-hand group of segments those that proceed in parallel to the wall but away from it), Lipkind was indeed successful in peeling off, as it were, wall segments from center segments (figure 13).

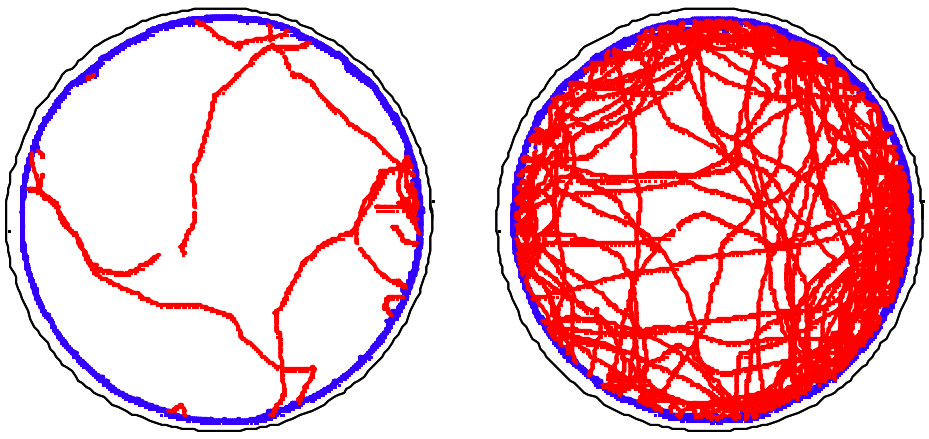

Figure 13

In computing the maximal (absolute) speed endpoint for each sub-population separately, she now got, across strains, highly significant differences for the wall segments (figure 14), and no difference for the center segments (not presented here).

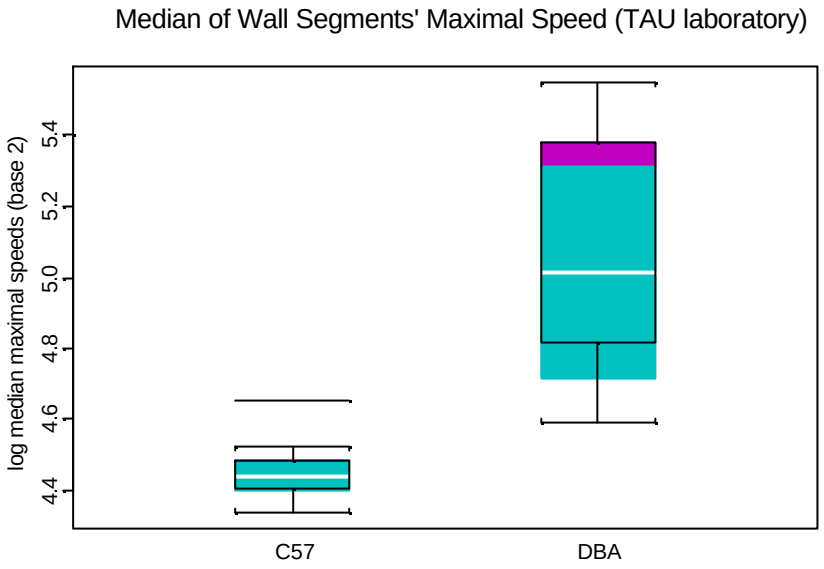

Figure 14

Had the study been performed only in Tel Aviv then that would have been the happy end of the study. However, as illustrated in figure 15, density graphs of radial speeds from NIDA and MPRC did not show a similar bimodality.

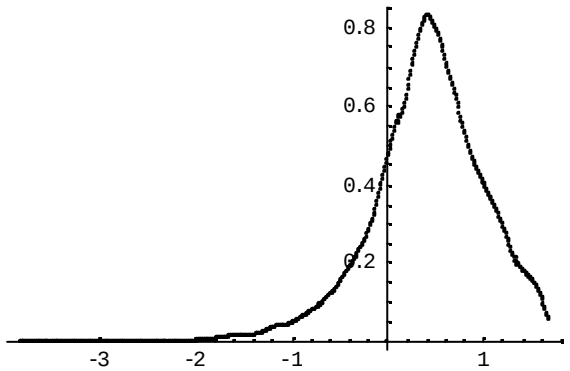

Figure 15: Density graph of log radial speeds from a NIDA Baltimore DBA mouse.

Examination of the PathPlots traced by the mice in these labs suggested that small imperfections in the circularity of the arena wall in these labs generated highr-than-zero radial speeds during travelling along the wall, and this in turn generated large enough noise to mask the distinctiveness of the wall segments group in the density graphs.
As illustrated in figure 16, the installation of new, fully circular, walls in these laboratories indeed uncovered the bimodality:

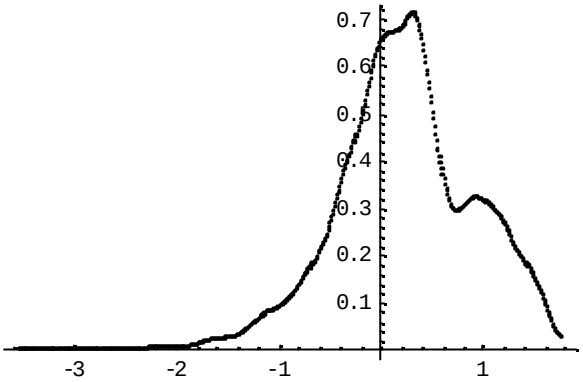

Figure 16: Density graph of log radial speeds from a NIDA Baltimore DBA mouse after fixing wall circularity.

The contribution of this process to the issue of standardization of experimental protocols across laboratories (see Crabbe et al., 1999; Wahlsten, 2001; Wurbel, 2000) is that *a priori* across the board standardization should perhaps be replaced by analysis-guided standardization.

An alternative solution to the imperfect circularity problem could be the use of software to estimate from the mouse path the actual form of the arena when it is not perfectly circular, and then smooth and repair the circularity.

A density graph of the radial velocities in the NIDA Baltimore mice after using SEE Arena Builder, a software developed by us to repair the circularity, illustrates how bimodaliy is uncovered after using this software (figure 17):

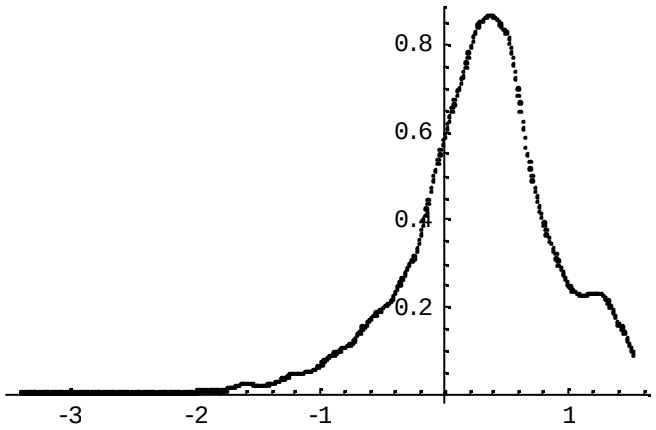

Figure 17

Our take away message from the success of this operation is that obsessive standardization may be replaced by carefully designed endpoints.

Having accomplished an intrinsic separation between wall and center segments in all 3 labs, between-strain comparison of maximal speed yields large strain differences in wall segments (figure 18), and no difference in center segments (figure 19):

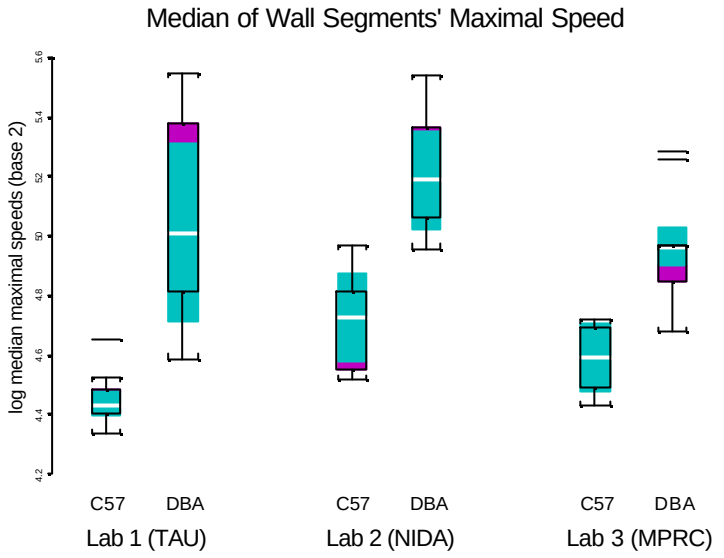

Figure 18

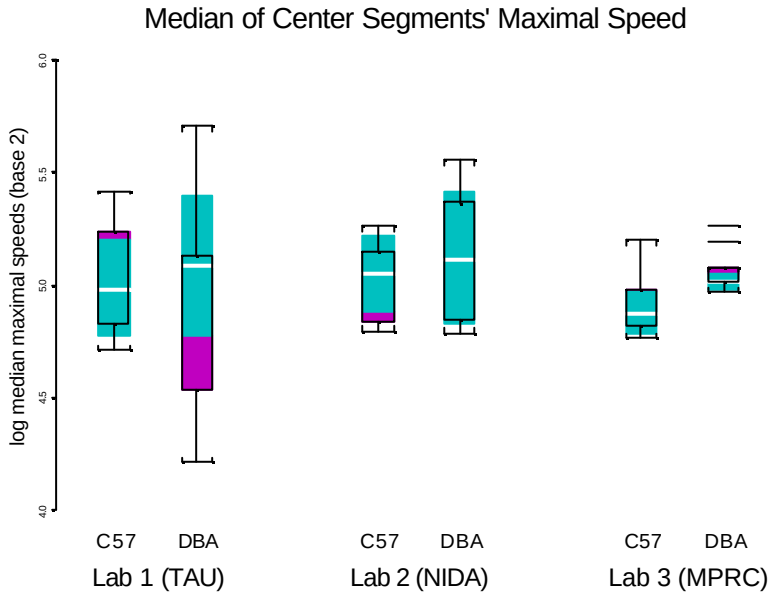

Figure 19

Second example: Computation of a compound endpoint called jitteriness (Neri Kafkafi, NIDA and MPRC Baltimore)

In our second example, Kafkafi started with a measure akin to segment acceleration which did not prove replicable, and ended up with an internally scaled measure he termed Jitteriness which proved both discriminative and replicable.

Figure 20 illustrates a time series of speeds including two progression segments (S1 and S2), and a lingering episode (LE), in between. The ratio between the

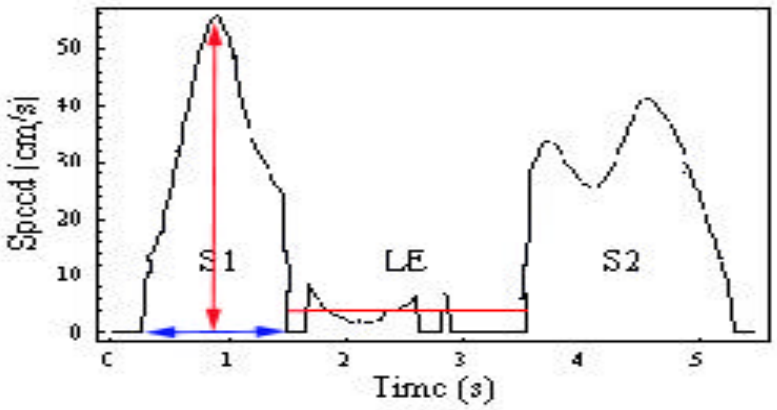

Figure 20.

maximal speed of a progression segment (red arrow in S1) and the segment's duration (blue arrow in S1) describes a measure akin to segment acceleration (peak speed divided by duration). Figure 21 describes the MSDR (Maximal Speed to Duration Ratio) values (mean SE) of C57 (black squares), and DBA (open diamonds) in 3 laboratories. As shown, there are large lab differences between strains for each strain separately (large site effects), and the two strains show different differences across laboratories (large interaction). In other words, the results are not replicable across laboratories (figure 21).

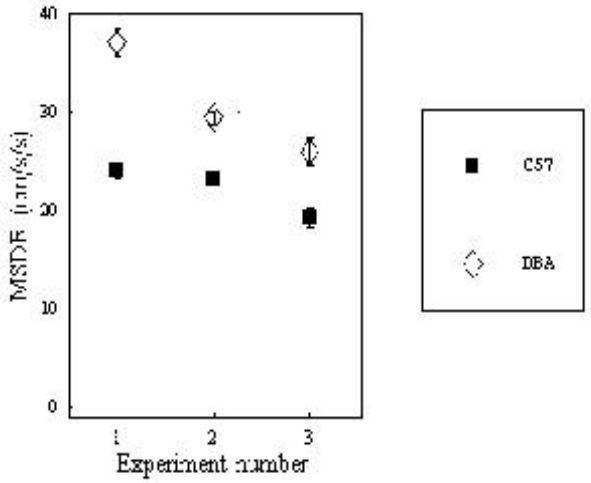

Figure 21.

While average lingering speed (see the red horizontal line in the segment LE in figure 20) yielded reciprocal results to those of MSDR values, In addition, Kafkafi had the impression that mice that show high accelerations also tend to freeze or show low lingering speeds. Therefore, he divided the MSDR values by average lingering speed, thus obtaining internal scaling for the new endpoint he called jitteriness. Figure 22 presents the Jitteriness (MSDR/average lingering speed) measure (mean SE) of C57 (black squares), and DBA (open diamonds) in the 3 labs. This time there are small lab differences for each strain separately (small site effects), and similar strain differences (similar vertical distance between strain values) across labs.

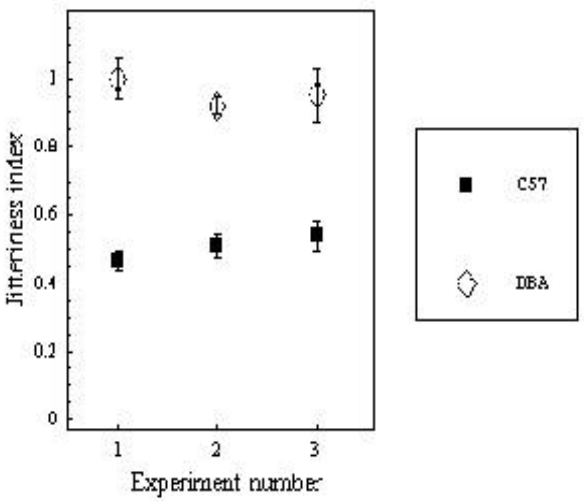

Figure 22.

The take away message from Kafkafi's study is that analysis-guided internal scaling might help design discriminative and replicable endpoints.

Multiple endpoints

We presently have some 32 carefully designed endpoints that are computed by SEE Endpoint Manager (SEM; figure 23), a stand alone user friendly publicly available software that calculates all these endpoints per data file. This and al the other programs can presently be obtained by writing to us at ilan99@post.tau.ac.il The software packages will all be placed on the web by April 2001.

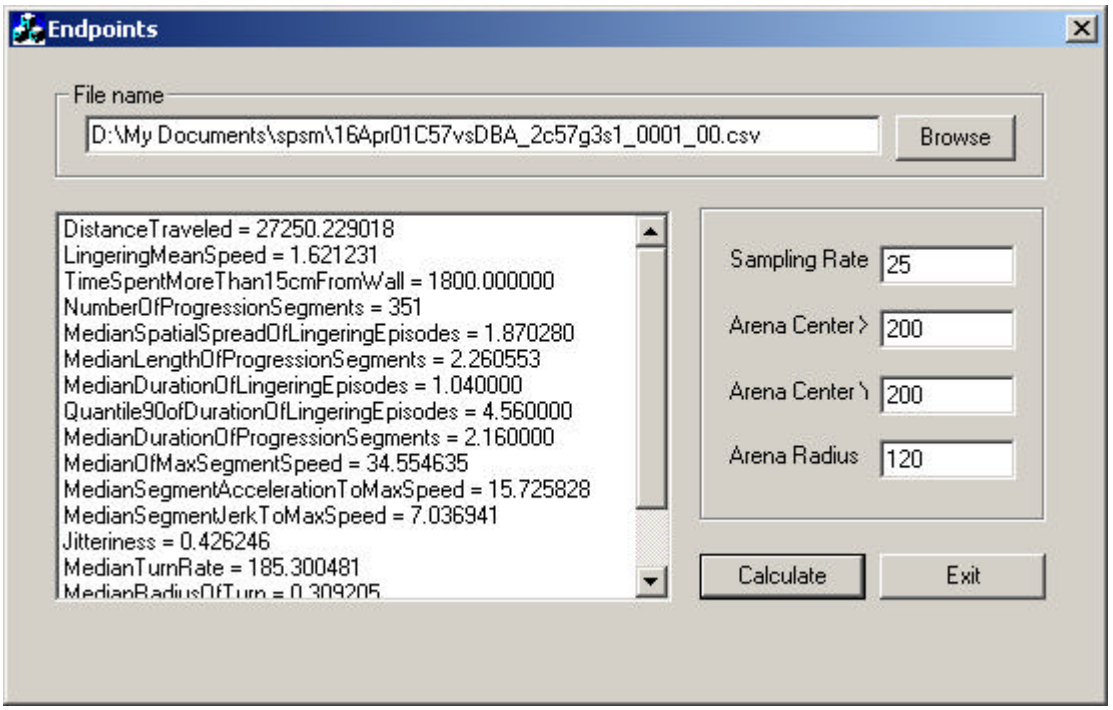

Figure 23.

Summary

For discrimination and replicability of locomotor behavior we use large space, relatively long time intervals, and high resolution.

After careful and robust smoothing, so as not to disrupt the fine structure of locomotor behavior, we cleave the data into intrinsically distinct classes of patterns and only then compute the endpoints for each class of patterns separately.

The demand for replicability while working in 3 laboratories simultaneously forced us to design measures of behavior that would not have been designed otherwise.

Acknowledgement:

This study is part of the project “Phenotyping mouse exploratory behavior” supported by NIH 1 R01 NS40234-01.

References

Crabbe JC, Wahlsten D, Dudek BC Genetics of mouse behavior: Interactions with laboratory environment. *Science* 1999; 284: (5420) 1670-1672

Drai D, Benjamini Y, Golani I (2000) Statistical discrimination of natural modes of motion in rat exploratory behavior. *J.of Neuro. Methods 96 (2000) 119-131*

Drai D, Golani I, SEE: (2001) A tool for the visualization and analysis of rodent exploratory behavior., *Neuroscience & Biobehavioral reviews* 25 409-426

Drai D, Kafkafi N, Benjamini Y, Elmer G, and Golani I (Submitted) An algorithmic approach to the phenotyping of mouse exploratory behavior: BALB/cJ versus C57BL/6J.

Hen I, Benjamini Y, Kafkafi N and Golani I (in preparation) Robust smoothing techniques for computing velocities and arrests in rodent locomotor behavior.

Kafkafi N et al., (In preparation) Jitteriness: a quantitative movement pattern designed for discrimination and replicability in mice behavior genetics and psychopharmacology research.

Lipkind D et al., (in preparation) Statistical discrimination between movement along the wall and in the center of the open field in mice: its utility in computing discriminative and replicable behavioral measures.

Wahlsten D (2001) Standardizing tests of mouse behavior: Reasons, recommendations and reality. Physiology and Behavior 73 695-704.

Wurbel H (2000) Genetics of behavior: the standardization fallacy. Nat Genet 26 263.