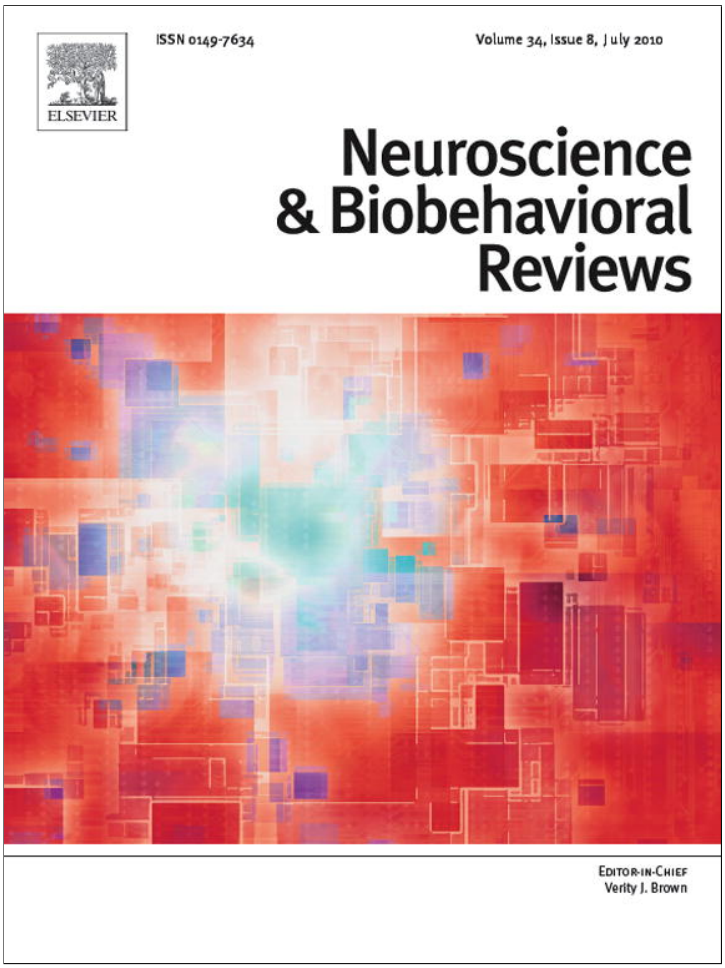

Neuroscience and Biobehavioral Reviews 34 (2010) 1351–1365

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

Ten ways to improve the quality of descriptions of whole-animal movement

Yoav Benjamini[a], Dina Lipkind[b], Guy Horev[b], Ehud Fonio[b], Neri Kafkafi[c], Ilan Golani[b,*]

[a] Department of Statistics and Operation Research, Sackler Faculty of Exact Sciences, Tel Aviv University, Israel
[b] Department of Zoology, Tel Aviv University, Wise Faculty of Life Sciences, Tel-Aviv 69978, Israel
[c] Maryland Psychiatric Research Centre, University of Maryland, Baltimore, MD, United States

ARTICLE INFO

Article history:
Received 19 April 2009
Received in revised form 11 April 2010
Accepted 12 April 2010

Keywords:
Segmentation of behavior
Phenotyping mouse behavior
Smoothing kinematic data
Compression of kinematic data
Replicability of results
Discriminability between strains and preparations
Genotype-laboratory interaction
Mixed-Model Anova
Open field behavior
Exploratory behavior
Description of behavior

ABSTRACT

The demand for replicability of behavioral results across laboratories is viewed as a burden in behavior genetics. We demonstrate how it can become an asset offering a quantitative criterion that guides the design of better ways to describe behavior. Passing the high benchmark dictated by the replicability demand requires less stressful and less restraining experimental setups, less noisy data, individually customized cutoff points between the building blocks of movement, and less variable yet discriminative dynamic representations that would capture more faithfully the nature of the behavior, unmasking similarities and differences and revealing novel animal-centered measures. Here we review ten tools that enhance replicability without compromising discrimination. While we demonstrate the usefulness of these tools in the context of inbred mouse exploratory behavior they can readily be used in any study involving a high-resolution analysis of spatial behavior. Viewing replicability as a design concept and using the ten methodological improvements may prove useful in many fields not necessarily related to spatial behavior.

Contents

* Corresponding author. Tel.: +972 3 6409391; fax: +972 3 6407350.
E-mail address: ilan99@post.tau.ac.il (I. Golani).

0149-7634/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neubiorev.2010.04.004

What is a good description? The suggestion that a "good" description of behavior is what a good ethologist considers to be a good description highlights a necessary but nonetheless insufficient condition for high-quality descriptions. This is because even good ethologists can err. How, then, if not on the basis of a claim to authority (Beer, 1980), is one to tell the difference between high and low quality descriptions?

The demand for replicability of results can help: One help comes from behavior genetics, where the genetic aspects underlying differences in behaviors are investigated. An example of an important tool in this endeavor is Quantitative Trait Loci (QTL) analysis, where particular regions on the chromosome are associated with quantitative traits that measure behavior in order to identify genes responsible for the change (Flint, 2003; Lewis, 2003). Using this method depends critically on the quality of the measurements of behavior. If, under the same standard conditions and for the same measure, in laboratory A one inbred mouse strain is found to be significantly higher than another inbred mouse, and in laboratory B the result is the opposite, then the locations found in laboratory A will not be replicated by laboratory B, making the results of both studies useless for gene localization studies. Such opposing results can mean that in at least one of the laboratories there may be an environmental factor, perhaps unknown to the experimenter, which is interacting with each of the strains in a different way. For example, the experimenter may not know that one of the strains is blind and that the lights are brighter in one of the laboratories, thus affecting only the sighted strain. This so-called *interaction* between laboratory and strain makes any comparison between the behavioral trait of the two strains questionable: if, for example, the aggressiveness of Bull Terrier dogs would be scored significantly higher than poodles in Stranger-directed Aggression in one clinic (as indeed they were; Blackshaw, 1991) and lower than poodles in another, then something would have been either wrong with (i) the experimental setup (inappropriate experimental setup and/or protocol), or with (ii) the way aggression was measured (a deficient measurement methodology), or, if we stick with the measure and protocol, there is something wrong with (iii) the way the significance was assessed (a deficient statistical methodology).

Standardization is necessary but not sufficient. A well accepted and widely discussed way to improve replicability is through standardization of the experimental protocol and measurement process. It is no wonder, therefore, that the Crabbe et al. multi-laboratory experiment (1999) had at the time a substantial impact on the behavior genetics community. In spite of their serious efforts to assure a high level of standardization, their study found significant genotype/laboratory interactions, thereby showing that some behavioral results were indeed idiosyncratic to a particular laboratory!

One response to this was a demand for ever more rigorous standardization (e.g., Beynen et al., 2001; van der Staay and Steckler, 2002). Our own group, advocating an ethological perspective, has suggested that while standardization is needed, it cannot completely solve the problem because it addresses the known factors, while interaction is often caused by unknown ones. We argue that variability across laboratories, like individual variability, is a fact of life and one that cannot be entirely eradicated. Therefore, improvement should be made in measurement technology. If we take this approach, the lack of replicability as measured by the Lab by Strain interaction can become a useful yardstick for guiding us to successful and therefore higher quality measures.

The need for a revised methodology of description: We argued that the essential inability to construct identical experimental setups, which was reflected in the Crabbe et al. (1999) results, can and should be used to filter out non-robust strain differences. Conversely, the magnitude of such non-robust differences can guide us in a revision of the way in which behavior is measured and in the way differences in behavior across laboratories are estimated. To demonstrate the practicality of our methodology we applied it to a comparison of exploratory behavior in eight mouse genotypes across three laboratories, and showed that replicable measures can in fact be established, even though the analysis that takes into considerations the different behavior of strains across laboratories requires larger observed strain differences for establishing significance. The statistical details are not important; most important from the standpoint of the present review is that the benchmark that the observed strain differences should pass in order to uncover significant genotype differences is set higher than usual when taking into consideration replicability across laboratories. Since this higher benchmark limits the power of discrimination of the system, a revision in the methodology of the measurement of behavior is implied: the higher benchmark requires measures that would be able to discriminate genetic differences even over the background of the interaction.

An improved measurement methodology would clearly not eliminate environmental influences, nor would it change the way laboratory variables interact with genetic factors. It can, however, reduce the effect of both on the measured behavior by taking into account some of the intrinsic properties of the specific animal, strain or laboratory. The development of an adequate measurement methodology is, therefore, not a luxury but a requirement dictated by the need for replicability of results across laboratories (Kafkafi et al., 2005).

What issues should a revised methodology of measurement address? One set of issues pertains to the use of behavior patterns versus the use of continuous variables: what features of movement justify scoring of discrete patterns and what features indicate a representation in terms of continuous kinematic variables? Should the continuous data time-series be compressed into a sequence of discrete whole-animal patterns? What type of patterns? At what stage of the analysis? Can compression be accomplished without recourse to subjective judgment, and without turning these patterns into black boxes whose kinematic content will become inaccessible? How can one select, or even design, variables that are more relevant than others in the face of a large number of candidate kinematic variables?

The second set of issues pertains to data preparation for analysis: to what extent does the output of a tracking system represent reality? What, if at all, should the criteria for segmentation of the flow be? Are there any costs to the segmentation process? How reversible should the segmentation be? How compatible are various segmentations of the same data time-series?

In the following review we describe ten ways to address these issues, focusing on how the degree of replicability and discriminability of results can be used as a beacon that would guide us toward better descriptions of behavior. The aim of this review is thus not to provide a review of better ways to describe behavior, but to show how better ways can be developed by assessing the quality of a description using objective criteria of replicability and discriminability. To achieve this aim we took the liberty of using examples from our own work and presented the material in a reader-friendly way, avoiding as much as possible technical aspects. This was accomplished by the use of multiple visualizations and graphs, and by attending to the statistical aspects of the replicability issue in a minimal way only where absolutely required.

1. Starting with the measurement of kinematic variables

1.1. Problems with the use of ad hoc whole-animal behavior patterns

A methodology of measurement that commences from a list of *ad hoc* whole-animal behavior patterns, such as the ethogram, is

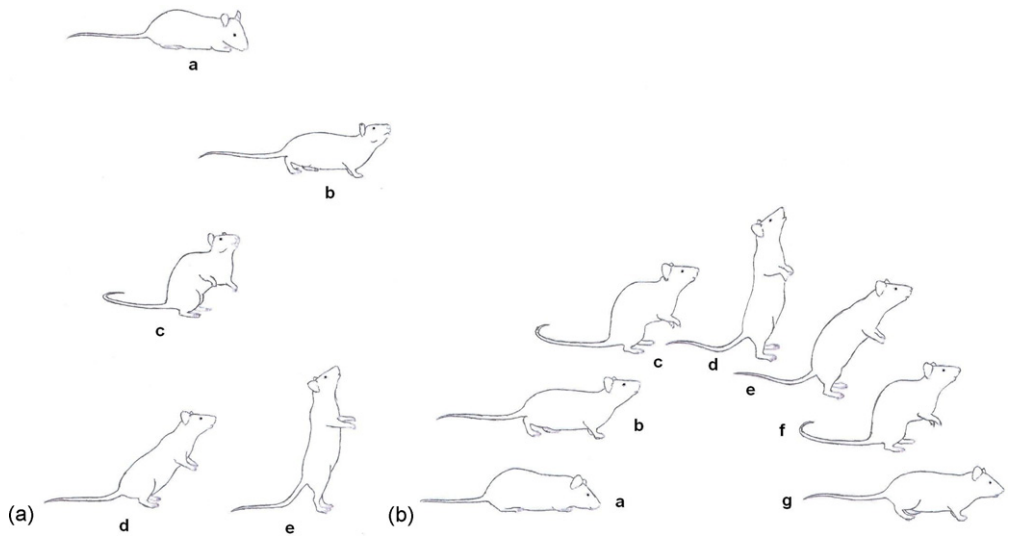

Fig. 1. (a) A selection of postures illustrating various forms of head raising behavior 'b–e' as well as no head raising 'a'. All (except 'a') could be scored as "rearing". (b) Successive stages of a single rearing activity (a–g; from: A. Neuman, M.Sc. thesis, Department of zoology, Tel Aviv University, 1990).

problematic. Such patterns could only be identified at the onset of a study and then used as the elementary building blocks of behavior, had they been composed of a fixed content. In the majority of cases, however, the content of these patterns is variable. Stopping episodes in mice, for example, typically consist of a variety of staying-in-place behaviors such as stepping in place, and horizontal and vertical scans that have variable durations, speed profiles and spatial spreads (Kafkafi et al., 2003a). These are not insignificant details; the scans disclose the direction of attention of the mouse, the extent of feet contact with the ground discloses the extent of the mouse's familiarity with the environment, etc. Similarly, even as straightforward a pattern as rearing behavior in rodents (Fig. 1a–e) is a hodge-podge category, encompassing versatile kinematic features such as the extent of hind feet contact with the substrate (e.g., plantigrade (c), versus digitigrade (b and e)), orientation of lower torso (e.g., horizontal (b), diagonal up (d), and vertical (c and e)), orientation of upper torso (e.g., horizontal (a–c), diagonal up (d), and vertical (e)), and orientation of head (e.g., horizontal (b–d) and diagonal up (e)). These features disclose information that could be critical with regard to the animal's momentary emotional state, the direction of the animal's attention and its intention.

In addition, premature categorization ignores the dynamics of movement taking place within the "pattern": pharmacological interventions, for example, may not affect the frequency of rearing yet alter the speed or sequence of postures involved in different rearing episodes (Fig. 1b). All this information is irrevocably lost once the behavior has been encapsulated in an *ad hoc* black box labeled "rearing". Neither inter-observer reliability nor a computational follow-up of this classification process by a neural network trained by a skilled human observer (e.g., Steele et al., 2007), can compensate for this information loss.

1.2. Advantages of the measurement of continuous variables

A methodology of measurement should thus commence with raw kinematic quantities (variables) of the movement material collected by tracking systems. This includes variables such as the location time-series and its derivatives (path scale), and the relations and changes of relation between the parts of the body (joints scale). Analysis has revealed that these variables can sometimes be divided into discrete segments. For example, segmentation of the speed time-series of individual mice and individual fruit flies (*Drosophila melanogaster*) may yield stopping episodes and progression segments (Drai et al., 2000; Valente et al., 2007; Branson et al., 2009). Similarly, recording the instantaneous position, speed, and turning rate of single worms (*Caenorhabditis elegans*) as a function of time in gradients of a chemical attractant, and subsequent analysis of turning rate yields segments of smooth swimming (runs) and episodes of frequent turning (pirouettes; Pierce-Shimomura et al., 1999). Discrete patterns like stopping and pirouetting thus constitute the *results* of the study, not its beginning. Both stopping and pirouetting are defined on the basis of a single kinematic variable, be it maximal speed or turning frequency. In this sense they constitute segment types rather than classical multidimensional behavior patterns.

Continuous kinematic measurements provide a quantitative detailed assessment of the changes taking place within segments. The details of how an animal performs a behavior (e.g., Fig. 1b) can be objectively quantified, highlighting the difference in the performance of the "same" segment type. The speed profile of a progression segment tells us whether a rat (*Rattus norvegicus*) "thinks" it is progressing away or toward its home base, or how familiar the immediate environment is. In an unfamiliar environment outbound speeds are higher than inbound speeds, while on a beaten path the speed ratio is reversed (Tchernichovski and Golani, 1995; Tchernichovski et al., 1998). Whishaw and colleagues used the speed profile in a well-trodden arena to show that unlike intact rats, hippocampal rats do not "know" where their home base is located (Wallace et al., 2002, 2008; Whishaw et al., 2001). Continuous simultaneous recording of location, of body orientation in reference to the wall and of speed reveals that mice speed up upon finding themselves at a distance from the wall and parallel to it, thus disclosing the fact that they use the wall as a guidance (Fig. 10), in some strains even when presumably blind (Horev et al., 2007). Finally, for many years, birdsong studies commenced from the developmental stage where syllables and songs could be identified and labeled in the sonagram. The preceding (subsong) stage, where syllable types could not be recognized by visual inspection, was mostly ignored. More recently, starting with a continuous recording of song development, and computing continuous features such as pitch, frequency modulation and entropy, it became possible to trace syllables back into the very beginning of the subsong developmental stage, where the morphogenesis of syllables and songs could be followed and deciphered (Tchernichovski et al., 2001; Tchernichovski and Mitra, 2002)."

The measurement of continuous variables has further advantages for segmentation, as detailed in Sections 4 and 6 below (Figs. 3 and 8).

2. Unfettering behavior from the constraints imposed by a small cage, short session and coercion

The larger the experimental environment and the longer the session the more stretched out the behavior is and the finer the spatial and temporal resolutions. This applies to location and its derivatives—speed, acceleration, curvature, heading direction, and higher level constructs such as the spatial spread of segments of behavior, preferred locations, and separation between edge and centre behavior. Discrimination between slow and fast strains, straight-path and curved-path walkers, wall-huggers and centre-dwellers improves in a large environment.

Finally, a large open space, combined with a free access to the arena from the animal's own home-cage has an extraordinary facilitating influence on the richness of the behavioral repertoire, which unfolds gradually, in a stable sequence, revealing the organization of this behavior (Fonio et al., 2009; see Section 8).

3. Robust data analysis at the preparatory stage

The use of kinematic variables implies an extensive preparation of the data for analysis, including robust (outliers resistant) smoothing. Noise and tracking artifacts are inherent to all tracking systems and critically affect the results. Even as straightforward a measure as Distance Traveled can be misleading. Whereas without proper smoothing an anaesthetized (and therefore still) mouse was found by us to "travel" in 15 min, 94 m with an average speed of 10 cm/s, after smoothing Distance Traveled was reduced to 6 m! The reason for this is that if the boundaries of the tracked animal do not fall exactly on one pixel, they might vacillate between two neighboring pixels, so that the "centre of gravity" (by whatever method it is calculated) will waver between discrete values even if the animal is stationary. The discretization of the measurement process affects not only a stationary, but also a moving mouse (see Fig. 1 in Hen et al., 2004). This example demonstrates how the smoothed data are better estimates of the trajectory of the mouse than the actual time-series of its recorded locations.

Since speed and acceleration disclose the "forces" acting on the animal, it is extremely important to obtain faithful records of these quantities in order to understand the interaction between the animal and its environment. While smoothing is a necessity, smoothing methods can have grave consequences on the outcome of a study: for example by eliminating short stops and thereby erroneously joining two movement segments into a single longer one; by shortening long stops; by flattening peak velocities; and by being influenced by outlier artifacts. In addition, the smoothing requirements of staying-in-place behavior differ from those of progression from one place to another. The magnitudes of the above-listed shortcomings are reduced by using robust nonparametric methods that are not affected by outlier artifacts, while the above-listed requirements are fulfilled by using a distinct smoothing method for each of the locomotor modes (Hen et al., 2004; Fig. 2a). The smoothing yields time-series of location data, speeds, heading directions, and, most important, it identifies arrests (Fig. 2b), which are subsequently used for the segmentation of the time-series into discrete behavioral units with proven ethological relevance (http://www.tau.ac.il/~ilan99/see/help/).

Noisy kinematic variables mask real similarities and real differences wherever they exist. Robust smoothing improves both replicability and discrimination by unmasking these similarities and differences.

4. Segmentation based on intrinsic geometrical and statistical properties of the kinematic material

In most open field studies cutoff points are established *ad hoc*: a stop, for example, is defined as staying in place for at least 0.5 s. It is, however, preferable to use "intrinsic" criteria for (i) finding out whether progressing and stopping do not lie on a continuum, and (ii) defining the cutoff point between them: "intrinsic" is not used here as a hand-wave. It rather means that the researcher does not impose on the data a predefined rule of what constitutes, e.g., a stop. To obtain an intrinsic measure an algorithm that processes jointly the entire time-series of the individual animal's measurements is defined. Cutoff points are then calculated based on statistical properties such as humps in the density or clusters in the joint distribution of measurements. In this way the cutoff reflects the individual animal's behavior, while still being algorithmically defined in an identical way in all animals under study. Not only does each strain define, by its own behavior, its own distinction between progression and staying in place—the cutoff point is customized for each individual, allowing one to determine how fixed this cutoff point is across animals, strains, and preparations (Drai et al., 2000). The individually customized cutoff points are used to establish a strain-specific measure that characterizes the strain (Kafkafi et al., 2005). Using individually customized cutoff points increases replicability by reducing the apparent variability brought about by *ad hoc* segmentation (Lipkind et al., 2004).

Fig. 2b presents a schematic illustration of the segmentation process as we describe below. At the first stage we isolate intervals of complete arrests, and the inter-arrest intervals in the speed time-series are defined as motion segments. The population of motion segments includes long-distance segments that reach high maximal speeds, as when a mouse moves from one place to another and very short-distance segments that attain very low maximal speed, as when a stationary mouse performs a lateral scan, or one to few steps, and then arrests. Plotting a frequency distribution of the maximal speeds of motion segments (Fig. 3) suggests that the population of motion segments performed by a mouse during a session in the open field indeed consists of a mixture of two distinct locomotor modes: high-speed and low-speed motion segments. At this stage the cutoff point between the two populations is obtained by fitting a Gaussian mixture model to the empirical density plot (Fig. 3). In particular, the Gaussian on the right consists of motion segments that reach high maximal speed, and carry the mouse from one place to another. These are the so-called progression segments. The last index of a progression segment and the first index of the next progression segment bound a so-called lingering episode (see Fig. 2b). Lingering episodes are thus packages that consist of at least one arrest and zero-to- several low-speed motion segments (e.g., the second lingering episode underlined by the second purple line in the bottom of Fig. 2b consists of three arrests and two low speed motion segments). The time-series data points are thus compressed automatically into a sequence of intrinsically defined segments. The individually customized cutoff point between lingering and progression segments was indeed found to contribute to the replicability of results across laboratories (Kafkafi et al., 2005).

4.1. How customized cutoff points increase replicability by reducing the interaction term

Let us return to the simple example where the experimenter may not know that one of the strains is blind, and that the lights are brighter in one of the laboratories, thus affecting only the sighted strain. Let us assume that this effect results in the sighted mice keeping closer to the wall, at, say, up to 5 cm, under the brighter-light situation, while the blind ones are not affected. Suppose that both strains at both labs spend the same amount of time near-

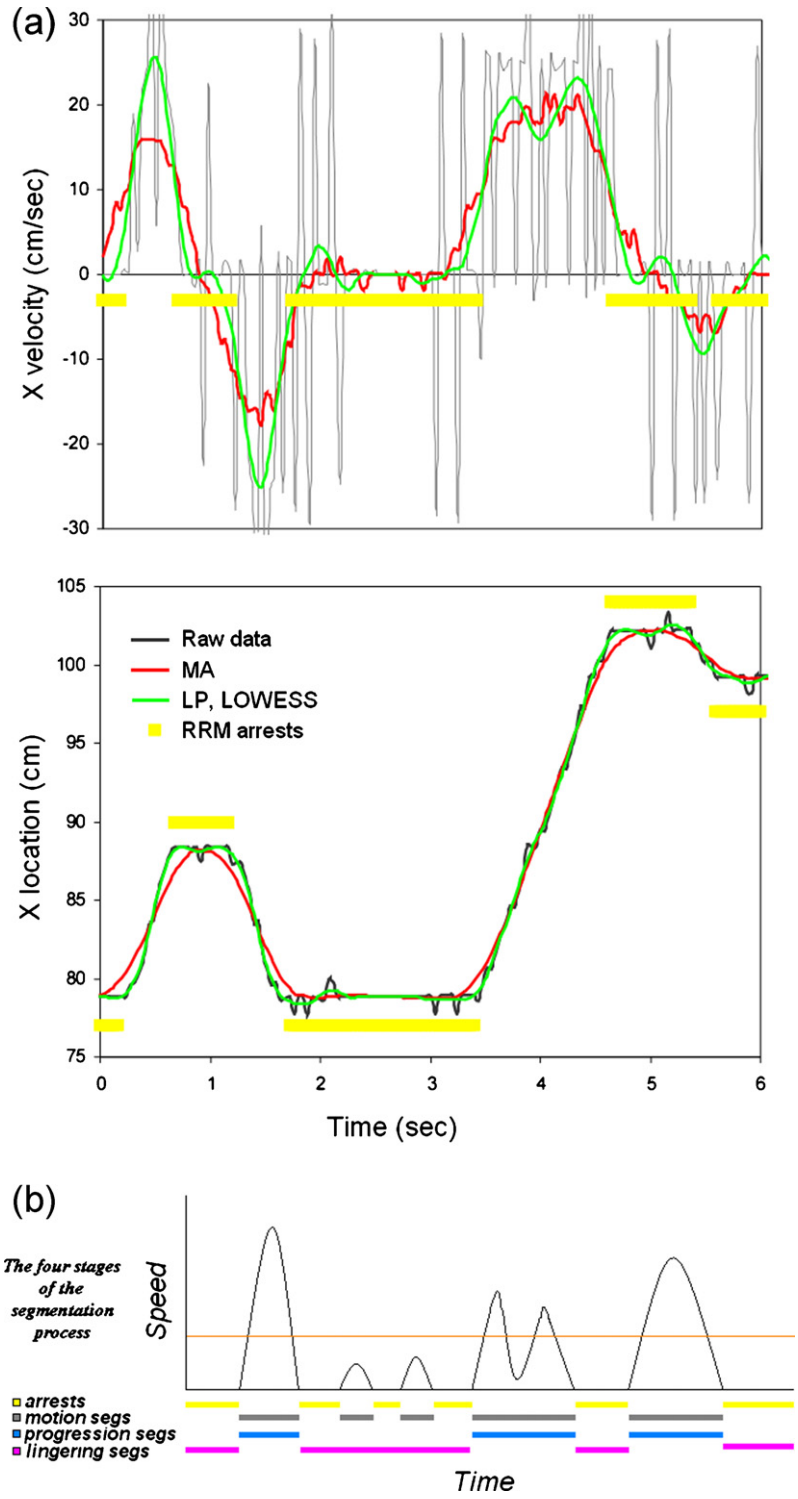

Fig. 2. (a) Location (bottom graph) and speed (top graph) out of 6 s time interval of a mouse's movement, smoothed using several methods. "Raw" speed (gray, top graph) is calculated as the differences between consecutive raw locations. MA (moving average) speed (top, red) is calculated as the differences between consecutive MA smoothed locations. LP (Local Polynomial) speed (top, green) is calculated directly by the LP. LOWESS-smoothed locations and velocities were almost identical to those calculated by LP since there were no outliers. Arrests are computed as zero speed in the RRM (Repeated Running Medians) smoothed series. The time ranges of arrests are denoted by yellow stripes (from Hen et al., 2004). (b) A schematic illustration of the four stages of the segmentation process: stage 1: the smoothing process yields intervals of complete arrest (six yellow segments in this illustration). Stage 2: inter-arrest intervals identify motion segments (five gray segments in this illustration). Stage 3: a density distribution of maximal speeds of motion segments yields a cutoff point (thin horizontal line parallel to the *X*-axis) that classifies the motion segments into 2 distinct maximal speed populations: low-speed motion segments that do not cross the cutoff speed, and high speed ones that do cross the cutoff threshold. Stage 4: these high-speed motion segments are defined as progression segments (three blue segments in this illustration). Stage 5: the intervals between progression segments are defined as lingering episodes (four purple segments in this illustration). As illustrated, a lingering segment may consist of one-to-several arrests and none to several low-speed motion segments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

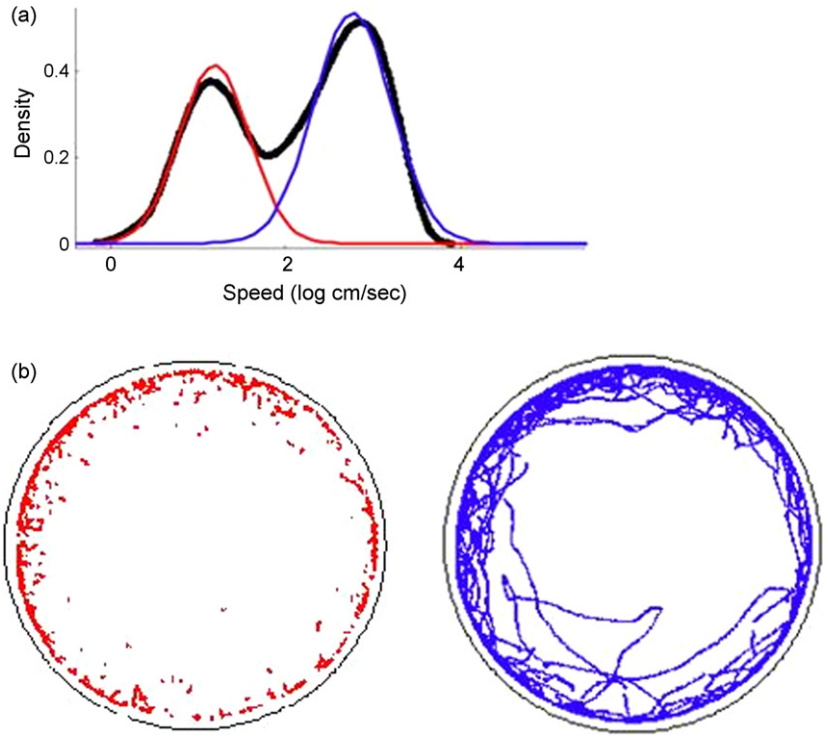

Fig. 3. (a) A density graph (a sliding window histogram) displaying the distribution of peak speeds of motion segments (black line) reveals two distinct populations: low-speed segments (lingering or stopping episodes) and high-speed segments (progression segments). Red and blue lines represent the two components of a Gaussian mixture model fitted by the EM (expectation–maximization) algorithm to the distribution, in order to establish a threshold that distinguishes between the two populations. (b) A visualization of the two separated populations: lingering episodes (red left) and progression segments (blue right) performed during the analyzed session (from Kafkafi et al., 2003a,b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the-wall. Still, if "near-the-wall" is defined *ad hoc* at, say, 10 cm, time spent near-the-wall will be higher for the sighted mice in the lighted lab as it will include also time spent away from wall. This will result in a lab effect (the average over the two strains will be higher for the lighted lab), as well as in a lab by strain interaction. If, instead, the size of the ring "near-the-wall" is allowed to be determined individually, both strains at both labs will spend equal time near-the-wall, reducing both lab effect and lab by strain interaction (Lipkind et al., 2004).

Using individually customized cutoff points between near-wall and centre segments uncovers a highly replicable and discriminative strain-specific measure characterizing the thickness of the ring of wall-cursions, which are segments that are performed near-the-wall and in parallel to it (Lipkind et al., 2004; Fig. 4).

The versatility of strain-specific wall-ring thickness had been fully masked by the commonly used *ad hoc* 10 cm threshold for near-wall behavior. Conversely, a description in terms of cus-

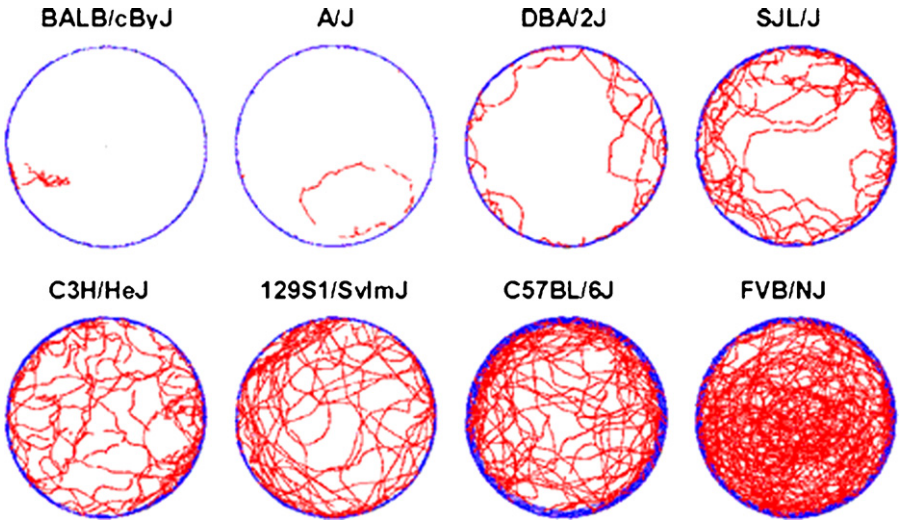

Fig. 4. Plots of the path traced during a 30 min Open field session by mice belonging to eight inbred strains. The path is separated into movement along the arena wall (blue wall-ring) and movement in the centre (red) using the Wall-Centre-Separation algorithm. Wall-ring width was found to be highly replicable across three laboratories (from Lipkind et al., 2004). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

tomized strain-specific wall-rings is preferable to a description based on an *ad hoc* distinction.

5. Segmentation at the indexing level leaves the data time-series transparent for further analysis

Compression of the stream of behavior into discrete patterns at the indexing level preserves the transparency of the segments so that their content can be subsequently analyzed, either in reference to the segments or regardless of them, thus highlighting other aspects of the organization of behavior. The obtained series of segments does not replace the original time-series of the kinematic variables. The original time-series are deposited along with their summaries in publically available data bases of behavior so that they can be used repeatedly for analysis from a variety of vantage points.

In our own work kinematic properties of segments are used to establish a variety of higher level behavioral constructs: the frequency and topographical distribution of lingering are used to define preferred places (Golani et al., 2005) while the probability of their performance at particular locations is used to define locational memory (Dvorkin et al., 2008). On the other hand, the speed profile within lingering episodes is used to calculate average lingering speed—a quantity characterizing the level of activity during staying-in-place behavior that was found to be highly heritable and discriminative across inbred strains (Kafkafi et al., 2003a). A segmentation based on the maximal speed attained within segments yields, for example, the sequence of progression segments and lingering episodes illustrated in Fig. 5a; and zooming into the content of the same segments in Fig. 5b provides access to the speed profile of these segments. It also provides access to other kinematic parameters including their acceleration-related properties (Fig. 5c) and their rate of turn (a measure akin to curvature), at one-to-several scales (Kafkafi and Elmer, 2005; Kafkafi et al., 2003b; Fig. 5d).

The multiplicity of measures gained by the time-series transparency, and therefore the accessibility of the data, increases the likelihood that at least some of the measures will pass the higher benchmark for replicability set by the incorporation of the interaction term into the statistical yardstick establishing significant differences between strains across laboratories. Conversely, the measures that passed the benchmark are better descriptors of the behavior.

6. Alternative complementary segmentations

The segmentation of the flow into progression segments and lingering episodes does not prevent alternative and complementary segmentations: such as into excursions performed from a home base (Fig. 6) – sequences of progression segments and lingering episodes that can in turn be divided into outbound and inbound portions (Tchernichovski et al., 1998): and incursions – forays into the centre that start and end at the wall (Fig. 7), which are in turn separated into discrete classes on the basis of their maximal distance from the wall, by fitting a Gaussian mixture model to an empirical density function (Fig. 8, Lipkind et al., 2004).

Once again, the multiplicity of segment types increases the likelihood that at least some of the measures derived from them will pass the benchmark for replicability and therefore provide better descriptors of the behavior.

7. Stratified measurement of segment types

When sub-populations vary considerably, it is advantageous to measure each subpopulation (stratum) independently. Stratifica-

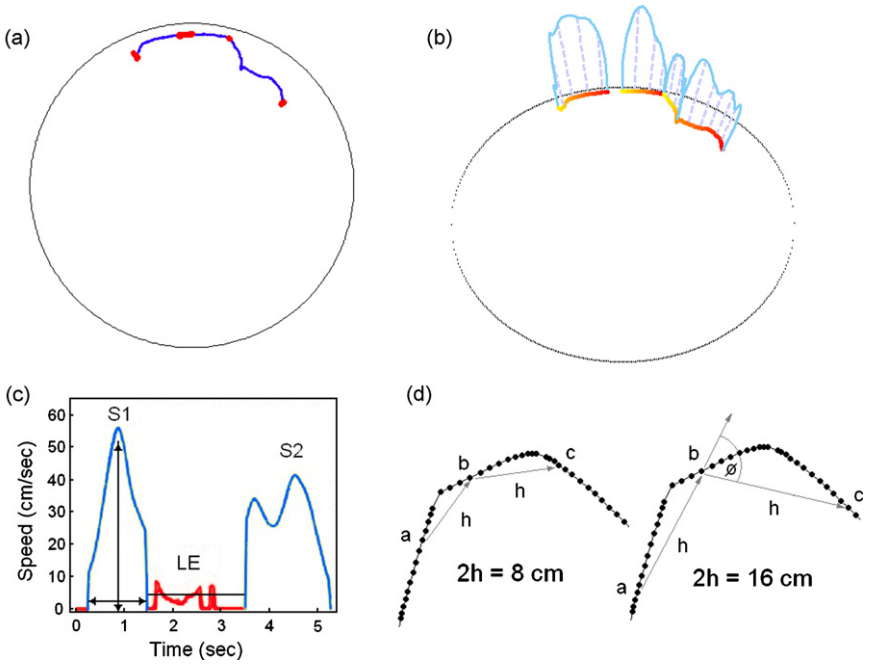

Fig. 5. (a) A portion of a mouse's path in the arena after smoothing and segmentation into progression segments (blue) and lingering episodes (red). (b) The same path visualized in a, but with the speed profile of the progression segments presented in the 3rd dimension (in azure). The direction of progression is marked by the color of the path, starting with yellow and proceeding to red in each progression segment. (c) A time-series of speeds including two progression segments (S1 and S2), and a lingering episode (LE) in between. The ratio between the maximal speed of a progression segment (vertical arrow in S1) and the segment's duration (horizontal arrow in S1) describes "Dart" (Kafkafi et al., 2003b)—a measure akin to segment acceleration (peak speed divided by duration). The horizontal line in LE designates median lingering speed (from Kafkafi et al., 2003a). (d) Computation of rate of turn: Examples of computing an endpoint related to curvature of the path at one data point, 'b', of a progression segment at two scales: 8-cm scale (left) and 16-cm scale (right). The curvature in degree/cm is defined as Ø/(2 h). The process is repeated at all data points belonging to progression segments and on several scales (from Kafkafi and Elmer, 2005). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

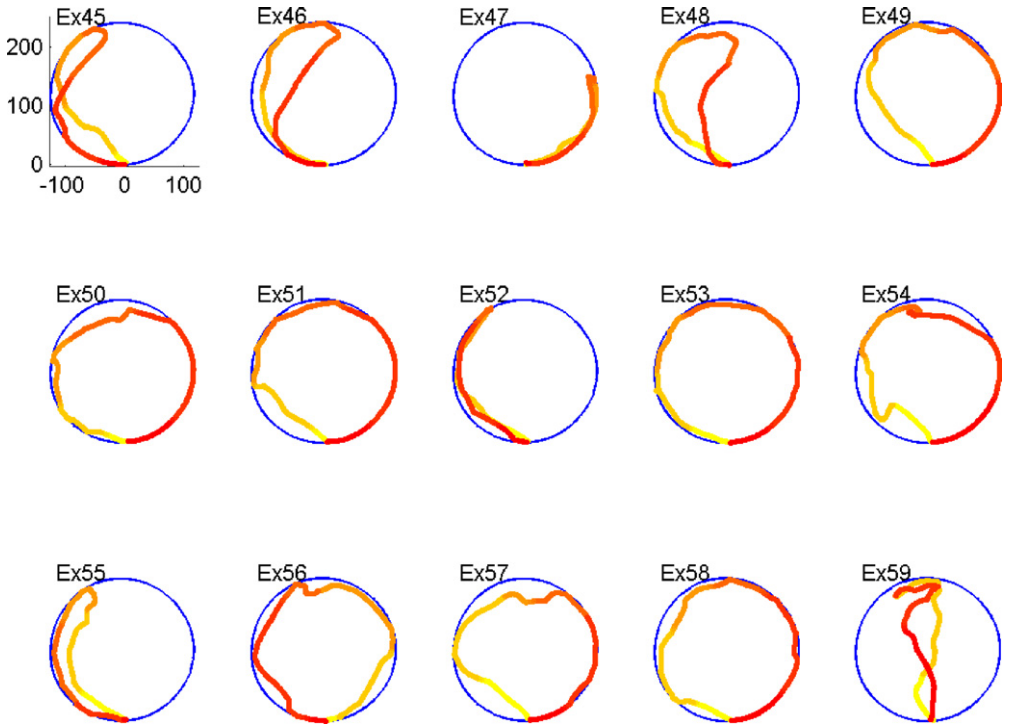

Fig. 6. Successive excursions (#45–59) in a free exploration setup from the attached home-cage into a 2.5 m diameter arena of a C57BL/6J mouse. Yellow to red signifies direction of progression (courtesy of Z. Havkin and D. Checkroun, unpublished results). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

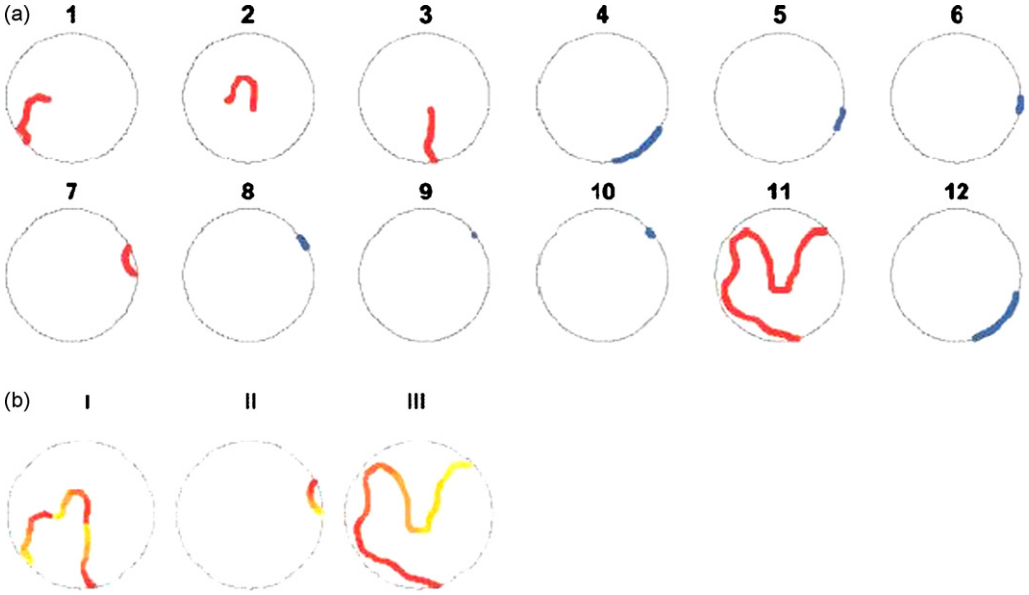

Fig. 7. (a) Plots of 12 successive segments of motion along the wall (blue) and in the centre (red) in a C57BL/6J mouse-session. (b) Plots of three successive incursions in the same session. Yellow to red coloring indicates the direction of movement in each segment within an incursion. Note that incursion I is composed of three centre segments (1, 2 and 3 in 'a'), and incursions II and III are each composed of a single centre segment (7 and 11 in 'a', respectively) (from Lipkind et al., 2004). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

tion is the process of grouping members of the population into relatively homogeneous subgroups before measurement.

7.1. Stratified measurement of segment types increases replicability

The dissection of the path into intrinsically defined segment types, followed by a measurement of the features of each of these types separately, yields more replicable endpoints than does a measurement of a mélange including some or all of them. A comparison of the numbers of Incursions performed during a session in, for example, C57BL/6J and DBA/2J mice shows that the number is higher in the first strain, but this difference is not statistically significant (Fig. 9 top left panel). Scoring of the three incursion types that are isolated by classifying incursions according to their maximal distance from wall (Fig. 8) is plotted in Fig. 9 top right and two bottom panels. The numbers of near-wall incursions are evidently not replicable across laboratories: although there is a large strain

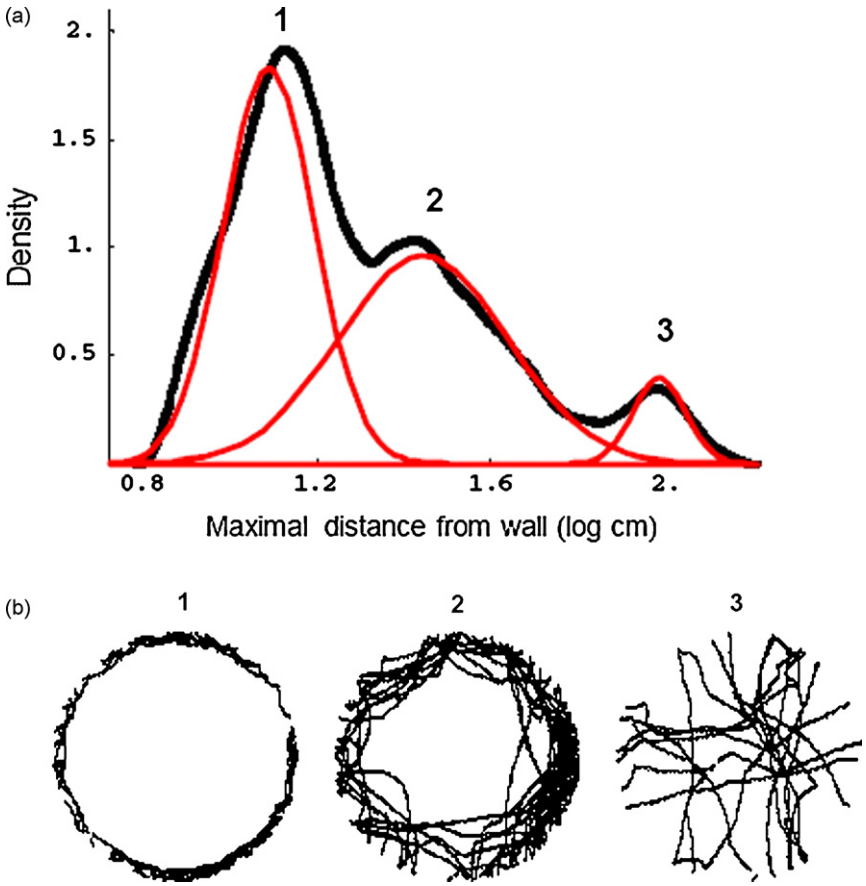

Fig. 8. (a) Black: a density graph of the distribution of the maximal distances from wall of centre segments (log transformed) in a single C57BL/6J session. Red: three Gaussians fitted to the distribution by the EM algorithm. The intersection points between the Gaussians serve as cutoff values for dividing all incursions performed in this session into three types. (b) Path plots of the incursions belonging to each type (from Lipkind et al., 2004). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

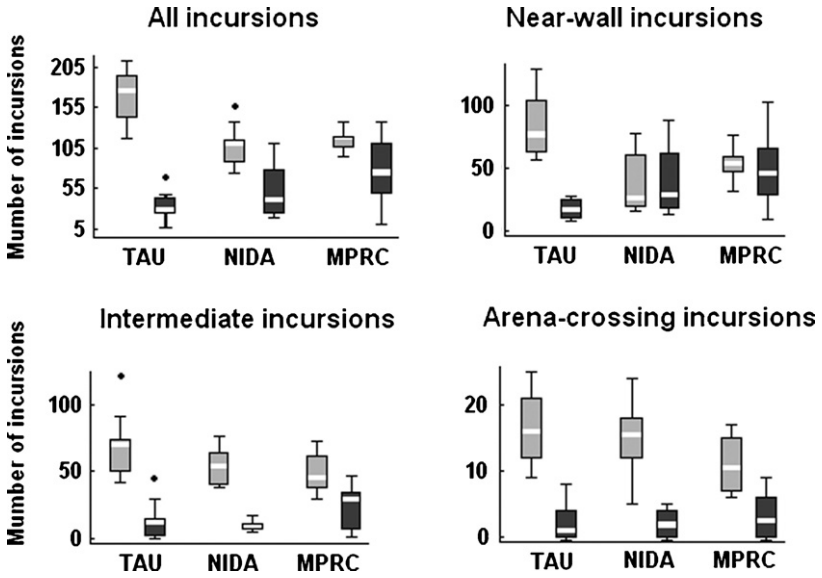

Fig. 9. Comparison of the degree of replicability of the number of incursions performed during a session in DBA/2J (in gray) and C57BL/6J mice (in black) before and after stratified measurement. Top left: measurement of all incursions pooled together (before stratification). The differences between the strains are not consistent (non-replicable) across laboratories. Top right and two bottom panels: After stratified measurement the difference between the strains across laboratories becomes more consistent in intermediate and in arena-crossing incursions. It is the non-replicable difference between near-wall incursions in the 2 strains that masked the highly replicable differences in the other 2 incursion types (from Lipkind et al., 2004).

difference in TAU, in the other two laboratories the two strains have a similar number of near-wall incursions. In contrast, there is a replicable strain difference in the number of intermediate and arena-crossing incursions, with C57BL/6J mice making significantly more incursions of both these types than DBA/2J mice in all three laboratories. It now becomes evident that the failure to achieve significant results in the overall number of incursions (Fig. 9 top left) is due to inter-laboratory variation in the numbers of a single incursion type: near-wall incursions. The boxplot summaries disclose how stratified scoring plus a multi-laboratory experiment transform low quality measures into high-quality measures: we start with non-significant differences in the un-stratified measure (Fig. 9 all incursions; partly overlapping spread of the boxplots of the two strains in all three laboratories), and continue with filtering out the segment type that shows a strong interaction (near-wall incursions), and we end up by keeping two new replicable measures, that of Number of Incursions in intermediate and that of Number of Incursions in arena-crossing incursions (that show a similar, consistent difference between the two strains' boxplots in all three laboratories).

7.2. Stratified analysis of segment types reveals complex interrelations between kinematic variables

The relatively erratic nature of mouse movement hinders a faithful computation of its heading direction and speed during lingering episodes. By sifting out the lingering episodes and computing these quantities only for progression segments, one can, however, obtain a faithful representation of the quantitative relations between the mouse's distance from wall, its heading direction and its speed (Fig. 10). As illustrated for two selected strains, each strain exhibits a strain-specific signature of the relationship between these quantities (Horev et al., 2007).

A selective depiction, only within progression segments of this threefold relationship, thus exposes the two types of influence exerted by the wall on the mouse's path: one of guidance and one of attraction. The guiding influence is expressed by the tendency of mice to progress in parallel to the wall. Although this tendency wanes with increasing distance from the wall, it can still be observed at large distances from it. The more parallel the mouse is to the wall, the higher is its speed, even when distant from the wall. This association between heading direction and speed shows that the mouse controls its heading in reference to the wall. This has also been observed in some blind strains, revealing that wall-guidance is not based exclusively on vision. The attraction influence is reflected by movement along the wall and by the asymmetry between speed during movement toward the wall, and speed during movement away from the wall: sighted mice move faster when moving toward the wall, whereas blind mice use similar speeds for both directions. Distinct influences of guidance and attraction have been shown to prevail in five inbred strains, revealing heritable components that were found to be replicable across three laboratories. This improved mode of description of open field behavior would not have been possible without the culling out for analysis of the progression segments stratum.

8. Design of animal-centered measures

Unthought-of measures and surprising building blocks of behavior are exposed by going back-and-forth between the videotaped behavior and its versatile visualizations. Sometimes the computational algorithms validate a hypothesis formed by unaided observation. For example, the distinction between lingering and progression segments was validated by plotting the density of maximal speed reached in motion segments (Fig. 3; Drai et al., 2000), or the discovery of "knots", a new type of places marked by the performance of tortuous paths full of twists and turns, was validated and characterized by the use of specially designed algorithms that scan the arena for places marked by high path curvature (Dvorkin et al., 2010). At other times it is the plot that sends the observer to the video. For example, plotting the number of stops per roundtrip yielded in rats an unexpected upper bound on that number. Counting on the video the stops made by a rat in the course of a roundtrip and being aware of this upper bound, an observer can asses the likelihood that the rat will turn around and rush all the way home without stopping (Golani et al., 1993). Be it as it may, both initial hypotheses and final validations are based on observation of the behavior on the video record.

Studying the moment-to-moment developmental dynamics of exploratory behavior in mice with our revised methodology yields 12 novel developmental landmarks and a host of mouse-centered measures characterizing their order of emergence and build up (Fonio et al., 2009; Fig. 11). The stability of their sequencing supports their validity as high-quality descriptors of the mice functional world (*umwelt*; von Uexküll, 1957).

9. Design of measures on the basis of data collected simultaneously in more than one laboratory

Instead of designing measures in one laboratory and only then validating them in subsequent studies performed in other laboratories, new measures are preferably designed from the outset on the basis of data obtained in parallel studies performed in several laboratories. Alternatively, the studies can be performed sequentially in more than one laboratory and final design decisions can be made in view of the results of all. Implementation of moderate standardization across laboratories increases the likelihood that the participating more-or-less similar laboratories represent a realistic sample of all laboratories (Kafkafi et al., 2005).

Our call for multi-laboratory experiments need not eliminate single-laboratory experiments. The user of a measure that has been already developed in more than one laboratory can conduct an experiment in a single laboratory, as long as the reported variances of laboratory interaction (see Section 10) are taken into account.

10. Set a higher yardstick using the Mixed-Model

Incorporation of the laboratory by strain interaction into the yardstick that determines the benchmark for replicability secures that the newly designed measures are indeed replicable.

Ten improvements. (i) Tracking of continuous kinematic variables, (ii) using ethological setups, (iii) robust smoothing, (iv) segmentation based on intrinsic properties of the kinematic data, (v) data transparency, (vi) alternative segmentations, (vii) stratified estimation of segment types, (viii) use of animal-centered measures that portray the animal's own functional world, and are (ix) designed and tested from the start across more than one laboratory and (x) incorporating into the statistical yardstick for replicability the strain/laboratory interaction, all contribute to the quality of a description.

The quality of a description is reflected in both the replicability and discriminability of the results. Extensive encapsulation of behavior into discrete "patterns" might increase the replicability of that behavior by concealing the variability of the kinematic content that is packed within the patterns. It would, however, also reduce the information content of that description, cutting in this way down its discriminative power. Conversely, a very detailed description of a behavior will contain a lot of information, yet a large proportion of that information might not be replicable. A good description would clearly optimize both replicability and discriminability.

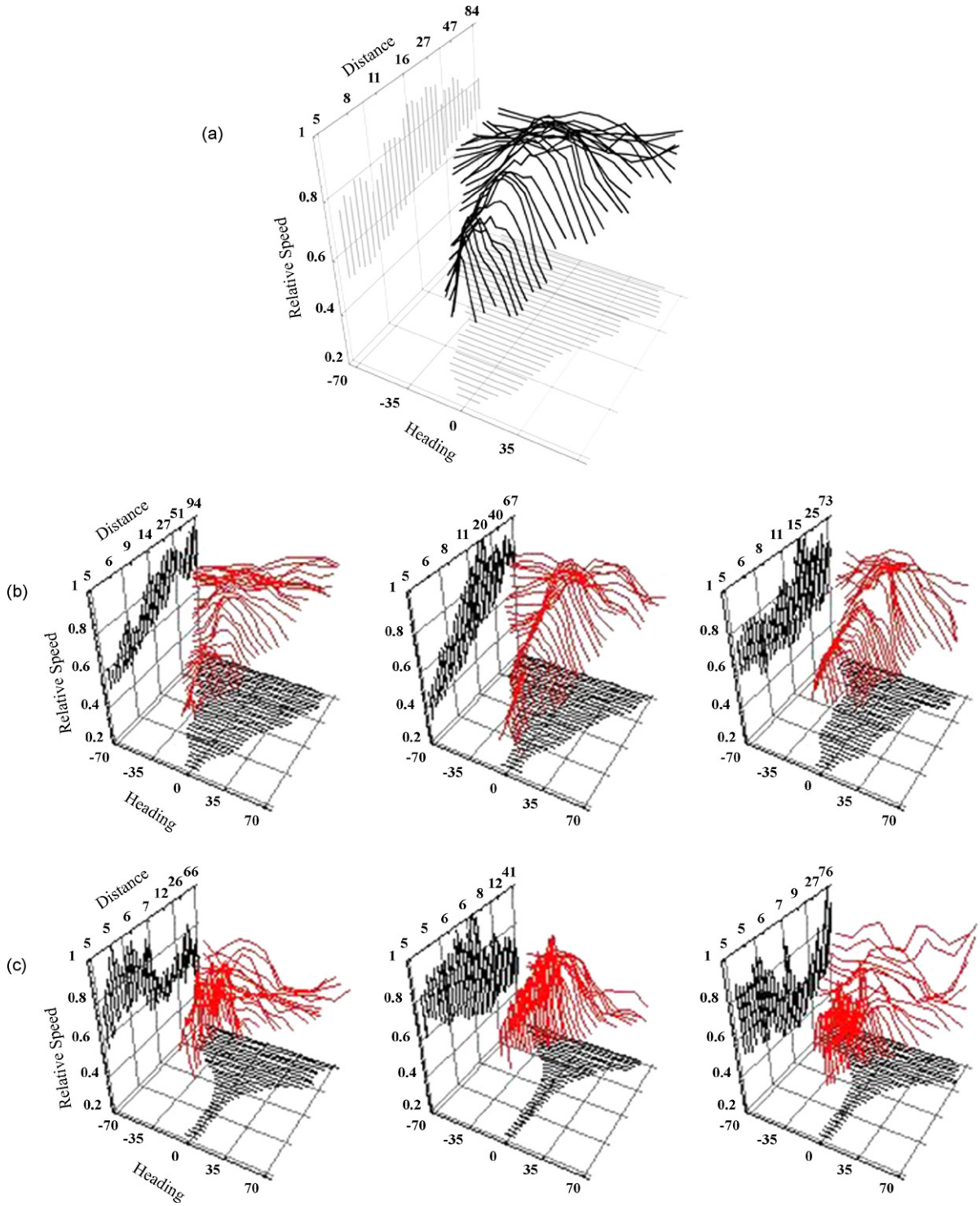

Fig. 10. Speed (*Z*-axis) plotted against heading (*X*-axis) and distance from wall (*Y*-axis) in representative sessions of C57BL/6J and SJL mice illustrate the threefold strain-specific relationship between these variables. (a) The three-dimensional landscape of the session of a selected C57BL/6J mouse is shown in black and its projection on the XY plane and the XZ plane are shown in gray. Heading values vary between −90° (outbound movement) and 90° (inbound movement) with 0° representing movement in parallel to the wall. The gradual increase of speed values depicted in the projection on the speed–distance plane shows the effect that distance has on speed. The gradual increase in heading ranges depicted in the projection on the heading–distance plane shows the effect that distance has on heading. The speed peaks at heading 0°, which form the central ridge in the three-dimensional graph, indicate the effect that heading has on speed. Note also that ridge steepness decreases with the increase in distance. This decrease points toward a second effect that distance has on speed, an effect that is mediated by heading. (b) Selected plots of 3 mouse-sessions of C57BL/6J mice illustrate the strain-specific signature of the speed–heading–distance relationship. The three-dimensional landscape is shown in red and its projection on the distance–heading plane and the distance–speed plane are shown in black. Note in all mouse-sessions the slight increase in speed values depicted in the projection on the distance–speed plane, the gradual increase in heading ranges depicted in the projection on the heading–distance plane, the speed peaks at heading 0°, which form the central ridge in the three-dimensional graph, and the decrease in ridge steepness with increasing distance. (c) Selected plots of 3 mouse-sessions of SJL mice illustrate their strain-specific signature. Note that the distance axis mostly extends over a much smaller distance-range compared to that in the C57BL/6J graphs. This is because each graph line represents an equal number of data points, and the SJL mice moved mostly near-the-wall. Note too the absence of speed augmentation with increasing distance from wall and the abrupt increase in heading ranges, as opposed to the gradual increase in heading ranges in C57BL/6J mice. Finally, note that the central ridge disappears in these strongly visually impaired mice at a distance of 6–7 cm from wall, as opposed to a distance of up to 50 cm in C57BL/6J mice (from Horev et al., 2007).

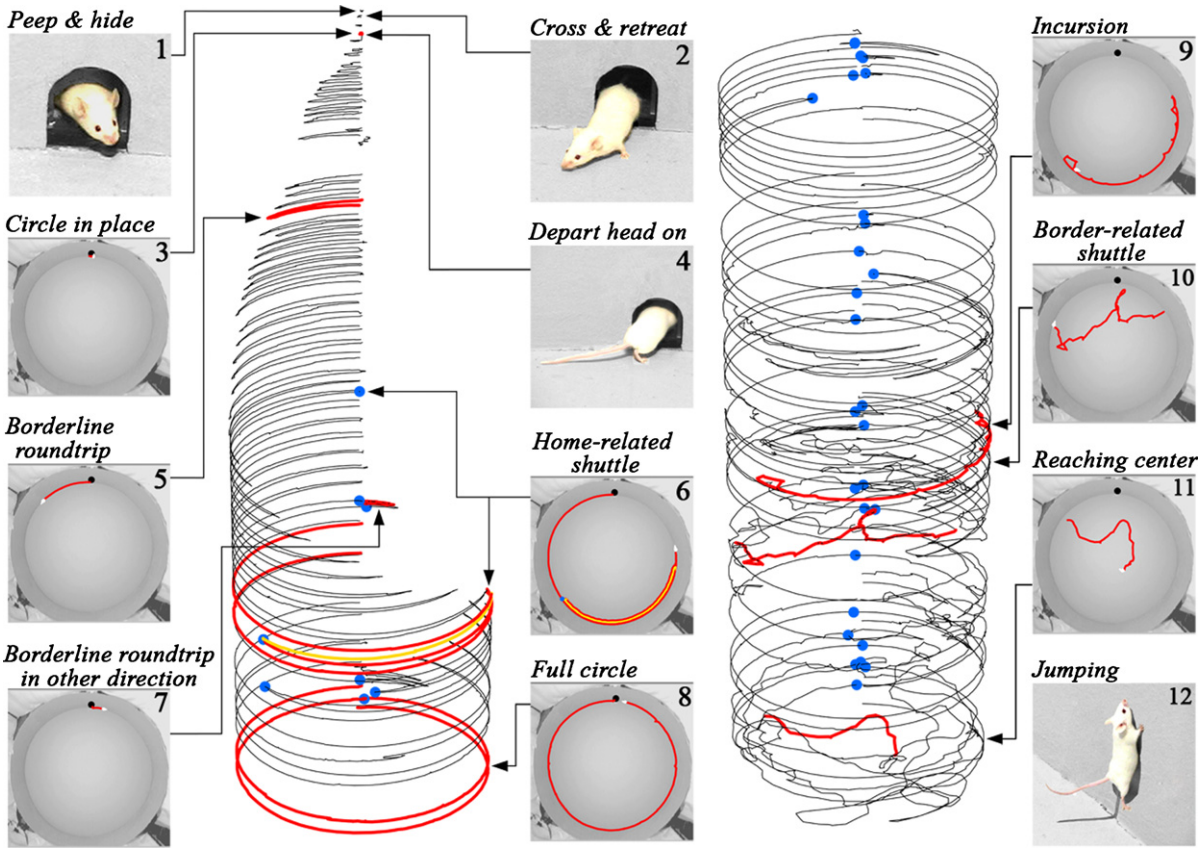

Fig. 11. The moment-to-moment developmental sequence of free exploration. The developmental landmarks in a specific BALB/c mouse-session performed across a 3 h period. The spiral proceeding from top to bottom, first in the left and then in the right column, presents the time-series of 2D locations on the path traced by the mouse. The enumerated figure-inserts show the 12 novel landmark motions exposed in this study, traced in red within the arena, and on the spiral. Blue dots indicate instances in which the mouse approached the cage doorway and did not enter the cage (cage-skips), or stopped short of returning all the way home during a return (home-related shuttle). Absence of a blue dot implies departure into home-cage. Yellow path stands for the return portion within a home-related shuttle (from Fonio et al., 2009).

(i) Replicability: Having implemented the ten improvements in the analysis of open-field behavior of eight genotypes across three laboratories using SEE (http://www.tau.ac.il/~ilan99/see/help/) it was demonstrated that replicable behavioral measures can indeed be practically established (Kafkafi et al., 2005).

The genotypic differences in all seventeen measures presented in Fig. 12 were highly significant when using the Mixed-Model ANOVA that incorporates the interaction into its statistical yardstick. These differences remained significant even after correcting for multiple comparisons (FDR-adjusted; Benjamini and Hochberg, 1995; Benjamini et al., 2001). Note, that according to the traditionally used Fixed-Model, more than half of the measures (marked by an asterisk in Fig. 12) would not have been found replicable. In contrast, use of the Mixed-Model reveals that the interaction term can be large, and yet discrimination would be good relative to that interaction. The revised methodology thus substantially increases replicability without compromising discriminability. Furthermore, the addition of many new mouse-centered measures increases the dimensionality of the space of behavioral traits, thereby increasing discriminability.

(ii) Discriminability: A SEE based comparison of 11 mouse strains, including two wild-derived inbred strains and first-generation-in-captivity wild mice, further demonstrates the contribution of the revised methodology to discriminability. Fig. 13 presents a visualization of the median values of twenty-seven calculated measures, by using a polygon-plot (a variation on the star-plot) for each of the strains. In this type of icon plot, the distance from the centre of the icon to consecutive corners of the polygon represents relative values of selected variables. Data points plotted in the centermost part represent low measure values, and data points plotted in the outermost part represent high values. The polygon thus highlights the unique multidimensional value characterizing a strain within the multidimensional space of open field behavior that is displayed by all strains; it also sensitizes the observer to the versatility of strain-specific profiles.

The thickness of the wall-ring traced by C57BL/6J mice (measure #1) contrasts with the thin wall-ring in DBA mice who actually brush the wall while progressing along it (see illustration in Fig. 4); the limited spatial spread of lingering episodes (measure #15) in Wild and wild-derived strains (Czechii and CAST), who hardly move around when staying in place, contrasts with the wide spatial spread brought about by stepping and scanning-around during lingering in DBA and SJL strains; and the apparently unexpected moderateness of all the values in the Wild mice contrasts with the extreme and often exaggerated values found in the domesticated *mus laboratorius* strains (Fonio et al., 2005).

The multiplicity of measures reflects the genuine complexity of locomotor and exploratory behavior. The multiple measures presented in Figs. 12 and 13 could be redundant. In particular, the two commonly used measures, Distance Traveled and Centre Time could explain most of the observed inter-strain variability, making the use of all the other measures excessive or even unnecessary. To examine the relative contribution of these two measures, Lipkind performed a Principal Component Analysis on 31 measures including these two, derived from movement material characterizing wall versus centre behavior in 8 inbred strains in the open field. In that study Distance Traveled and Centre Time explained only 9.6% of path variability

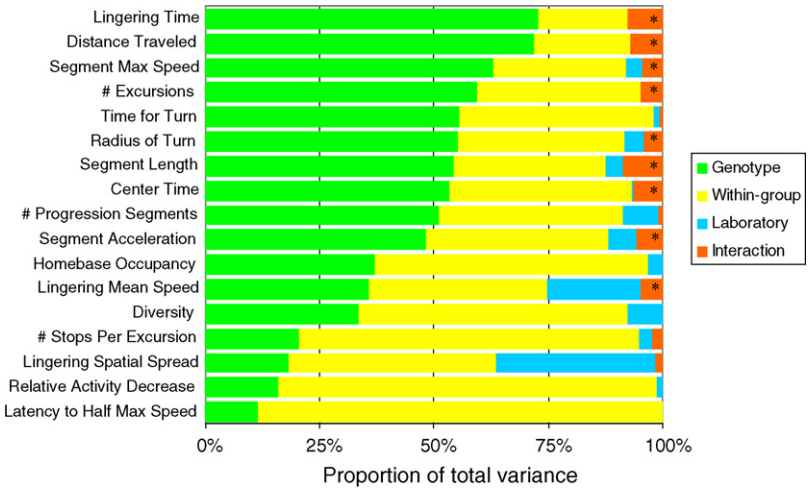

Fig. 12. The proportion of variance attributed to each of the factors—genotype, individual, laboratory, and interaction is shown by using the Mixed-Model. Measures are sorted by their proportion of genotypic variance. Genotypic differences for all endpoints are significant when using the Mixed-Model. Asterisks indicate interaction effects that were found to be significant according to the traditionally used Fixed-Model at a level of 5% (from Kafkafi et al., 2005).

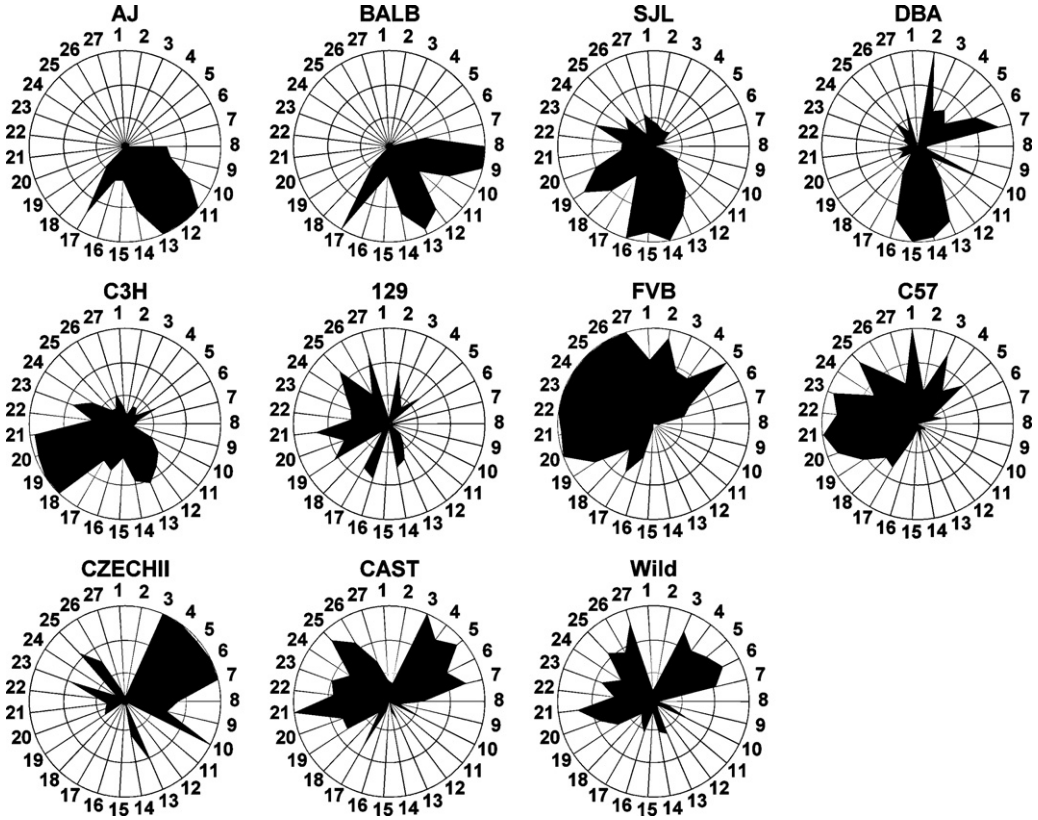

Fig. 13. Median values of 27 behavioral measures are visualized using a polygon-plot for each of 11 mouse strains (for list of measures see Appendix Table 1). A separate Polygon icon is plotted for each strain; relative values of the selected variables for each strain are represented by the distance from the Centre of the icon to consecutive corners of the polygon, with minimal values in the centre and maximal values on the perimeter. Variables are numbered, and presented in a clockwise order, in correspondence with the various measures (from Fonio et al., 2005).

across strains. The additional 29 variables were thus responsible for 90.4% of the overall variability, implying that the vast majority of information about path structure is not conveyed by the commonly used measures but by the newly developed ones (Lipkind, PhD thesis). Similarly, Horev, having demonstrated that the wall exerts both attraction and guidance influences on a mouse exploring an open field arena (Fig. 10) showed that these two influences, manifested by distinct measures, are independent of Centre Time (and therefore provide additional information). Guidance and attraction were furthermore shown to be relatively independent of each other (implying that they are mediated by different mechanisms; Horev et al., 2007).

It should furthermore be noted that when a measure is, for example, highly correlated with Distance Traveled in all but one strain, then using that measure is justified because it discriminates this strain from all the others and it highlights this measure as a distinct parameter shaping locomotor behavior. Furthermore, given a genetic difference or a pharmacological intervention, if it

affects directly one particular aspect of locomotor behavior that is somewhat correlated with, e.g., Distance Traveled, then this in itself would justify the separate monitoring of that aspect, because the effect would be smaller, more difficult to detect, and more difficult to interpret through its indirect influence on Distance Traveled. In summary, while a correlation between the multiple measures computed in SEE (Drai and Golani, 2001) would not obviate their separate measurement, we failed to establish such correlation. The revised methodology reviewed in this study reveals that locomotor behavior has a complex, highly heritable multidimensional structure that cannot be reduced to two or three dimensions without loosing most of the information contained in it.

Implications. Most or all of the ten recommendations can readily be applied to any study involving a high-resolution analysis of spatial behavior, be it the Elevated Plus Maze (Pellow et al., 1985), other types of mazes (e.g., the radial arm maze, Olton and Samuelson, 1976) including task related mazes such as the Morris water maze (Morris, 1984), and GPS monitored spatial behavior of wild animals in the field. The studied path may be traced by any organism, be it a fly (Branson et al., 2009; Valente et al., 2007), a human baby (Vitelson, 2005) or the eyes of a human subject exploring a landscape (Hayhoe and Ballard, 2005). All studies of spatial behavior will benefit from dynamic representations of location and its derivatives, from appropriate segmentation procedures based on intrinsic properties of the movement material, and from organism-centred variables and data transparency.

Generally speaking, viewing replicability as a design concept rather than a hurdle will prove useful in many other areas not necessarily related to spatial behavior; using intrinsic features of the data for classification, segmentation, stratification and establishment of organism-centred variables will yield individually adapted yet well-defined measurements in many fields; and even the finding of discrete patterns will be accompanied by retaining the quantitative properties of each pattern, when performed, allowing the dynamic study of its unfolding.

Acknowledgements

This study was supported by the Israel Science Foundation (ISF) of the Israeli Academy of Science Grant 915/05 (to I.G. and Y.B.). We thank Noldus Information Technology for the use of their EthoVision system including the new EthoVision XT 7.0. Initial versions of this paper were presented in a symposium organized by JC Fentress as part of a Measuring Behavior Conference held in Maastricht, The Netherlands, 2008, and in the Meeting of the Eastern Mediterranean Region of the International Biometric Society, Istanbul 2009.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neubiorev.2010.04.004.

References

Beer, C.G., 1980. Perspectives in animal behavior comparisons. In: Bornstein, M.H. (Ed.), Comparative Methods in Psychology. Erlbaum.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B Methodological 57 (1), 289–300.
Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., Golani, I., 2001. Controlling the false discovery rate in behavior genetics research. Behavioural Brain Research 125 (1/2), 279–284.
Beynen, A.C., Gärtner, K., van Zutphen, L.F.M., 2001. Standardization of animal experimentation. In: van Zutphen, L.F.M., Baumans, V., Beynen, A.C. (Eds.), Principles of Laboratory Animal Science, revised edition. Elsevier, Amsterdam, pp. 103–110.
Blackshaw, J.K., 1991. An overview of types of aggressive behaviour in dogs and methods of treatment. Applied Animal Behaviour Science 30, 351–354.
Branson, K., Robie, A.A., Bender, J., Perona, P., Dickinson, M.H., 2009. High-throughput ethomics in large groups of *Drosophila*. Nature Methods 6, 451–457.
Crabbe, J.C., Wahlsten, D., Dudek, B.C., 1999. Genetics of mouse behavior: interactions with lab environment. Science 284, 1670–1672.
Drai, D., Golani, I., 2001. SEE: a tool for the visualization and analysis of rodent exploratory behavior. Neuroscience & Biobehavioral Reviews 25 (July (5)), 409–426.
Drai, D., Benjamini, Y., Golani, I., 2000. Statistical discrimination of natural modes of motion in rat exploratory behavior. Journal of Neuroscience Methods 96 (2).
Dvorkin, A., Benjamini, Y., Golani, I., 2008. Mouse cognition-related behavior in the open-field: emergence of places of attraction. PLoS Computational Biology 4 (2), e1000027.
Dvorkin, A., Szechtman, H., Golani, I., 2010. Knots: attractive places with high path tortuosity in mouse open field exploration. PLoS Computational Biology 6 (1).
Flint, J., 2003. Analysis of quantitative trait loci that influence animal behavior. Journal of Neurobiology, published on line http://www.well.ox.ac.uk/flint/publications/JNeurobiol.pdf.
Fonio, E., Benjamini, Y., Sakov, A., Golani, I., 2005. Wild mouse open field behavior is embedded within the multidimensional data space spanned by laboratory inbred strains. Genes, Brain and Behavior 5 (5), 380–388.
Fonio, E., Benjamini, Y., Golani, I., 2009. Freedom of movement and the stability of its unfolding in free exploration in mice. PNAS 106 (50), 21335–21340.
Golani, I., Eilam, D., Benjamini, Y., 1993. Stopping behavior: constraints on exploration in rats (*Rattus norvegicus*). Behavioural Brain Research 53 (1/2), 21–33.
Golani, I., Benjamini, Y., Dvorkin, A., Lipkind, D., Kafkafi, N., 2005. Locomotor and exploratory behavior. In: Whishaw, I.Q., Kolb, B. (Eds.), The Behaviour of the Laboratory rat: A Handbook with Tests. Oxford University Press.
Hayhoe, M., Ballard, D., 2005. Eye movements in natural behavior. Trends in Cognitive Sciences 9 (4), 188–194.
Hen, I., Sakov, A., Kafkafi, N., Golani, I., Benjamini, Y., 2004. Journal of Neuroscience Methods 133 (1/2), 161–172.
Horev, G., Benjamini, Y., Sakov, A., Golani, I., 2007. Estimating wall guidance and attraction in mouse free locomotor behavior. Genes, Brain and Behavior 6 (1), 30–41.
Kafkafi, N., Elmer, G.I., 2005. Texture of locomotor path: a replicable characterization of a complex behavioral phenotype. Genes, Brain and Behavior 4 (7), 431–443.
Kafkafi, N., Lipkind, D., Benjamini, Y., Mayo, C.L., Elmer, G.I., Golani, I., 2003a. SEE locomotor behavior test discriminates C57BL/6J and DBA/2J mouse inbred strains across laboratories and protocol conditions. Behavioral Neuroscience 117 (3), 464–477.
Kafkafi, N., Pagis, M., Lipkind, D., Mayo, C.L., Benjamini, Y., Golani, I., Elmer, G.I., 2003b. Darting behavior: a quantitative movement pattern designed for discrimination and replicability in mouse locomotor behavior. Behavioral Brain Research 142 (1/2), 193–205.
Kafkafi, N., Benjamini, Y., Sakov, A., Elmer, G.I., Golani, I., 2005. Genotype–environment interactions in mouse behavior: a way out of the problem. Proceedings of the National Academy of Science of the United States of America 102 (12), 4619–4624.
Lewis, R., 2003. Multifactorial Traits. McGraw-Hill Higher Education.
Lipkind, D., Sakov, A., Kafkafi, N., Elmer, G.I., Benjamini, Y., Golani, I., 2004. New replicable anxiety-related measures of wall vs. centre behavior of mice in the open field test. Journal of Applied Physiology 97 (1), 347–359.
Morris, R., 1984. Developments of a water-maze procedure for studying spatial learning in the rat. Journal of Neuroscience Methods 11 (1), 47–60.
Olton, D.S., Samuelson, R.J., 1976. Remembrance of places passed: spatial memory in rats. Journal of Experimental Psychology: Animal Behavior Processes 2, 97–116.
Pellow, S., Chopin, P., File, S.E., Briley, M., 1985. Validation of open:closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. Journal of Neuroscience Methods 14, 149–167.
Pierce-Shimomura, J.T., Morse, T.M., Lockery, S.R., 1999. The Fundamental Role of Pirouettes in Caenorhabditis elegans Chemotaxis. The Journal of Neuroscience 19 (November (21)), 9557–9569.
Steele, A.D., Jackson, W.S., King, O.D., Lindquist, S., 2007. The power of automated high-resolution behavior analysis revealed by its application to mouse models of Huntington's and prion diseases. Proceedings of the National Academy of Science 104 (6), 1983–1988.
Tchernichovski, O., Golani, I., 1995. A phase plane representation of rat exploratory behavior. Journal of Neuroscience Methods 62 (1–2), 21–27.
Tchernichovski, O., Mitra, P.P., 2002. Towards quantification of vocal imitation in the zebra finch. Journal of Comparative Pathology A 188, 867–878.
Tchernichovski, O., Benjamini, Y., Golani, I., 1998. The dynamics of long-term exploration in the rat. Part I: a phase-plane analysis of the relationship between location and velocity. Biological Cybernetics 78 (6), 423–432.
Tchernichovski, O., Mitra, P.P., Lints, T., Nottebohm, F., 2001. Dynamics of the vocal imitation process: how a zebra finch learns its song. Science 291, 2564–2569.
von Uexküll, J., 1957. A Stroll through the worlds of animals and men: a picture book of invisible worlds. In: Schiller, C.H. (Ed.), Instinctive Behavior: The Development of a Modern Concept. International Universities Press, New York, pp. 5–80.
Valente, D., Golani, I., Mitra, P.P., 2007. Analysis of the trajectory of *Drosophila melanogaster* in a circular open field arena. PLoS ONE 2 (10).
van der Staay, F.J., Steckler, T., 2002. The fallacy of behavioral phenotyping without standardisation. Genes, Brain and Behavior 1, 9–13, Blackwell Munksgaard.
Vitelson, H., 2005. Spatial behavior in pre-walking infants: patterns of locomotion in a novel environment. PhD thesis submitted to the senate of Tel Aviv University.

Wallace, D.G., Hines, D.J., Pellis, S.M., Whishaw, I.Q., 2002. Vestibular Information Is Required for Dead Reckoning in the Rat. The Journal of Neuroscience 22 (22), 10009–10017.

Wallace, D.G., Martin, M.M., Winter, S.S., 2008. Fractionating dead reckoning: role of the compass, odometer, logbook, and home base establishment in spatial orientation. Naturwissenschaften 95 (11), 1011–1026.

Whishaw, I.Q., Hines, D.J., Wallace, D.G., 2001. Dead reckoning (path integration) requires the hippocampal formation: evidence from spontaneous exploration and spatial learning tasks in light (allothetic) and dark (idiothetic) tests. Behavioural Brain Research 127 (1-2), 14, 49-69.