

A novel method for automatic quantification of psychostimulant-evoked route-tracing stereotypy: application to *Mus musculus*

Stephen J. Bonasera · A. Katrin Schenk ·
Evan J. Luxenberg · Laurence H. Tecott

Received: 23 April 2007 / Accepted: 17 October 2007 / Published online: 21 December 2007
© Springer-Verlag 2007

Abstract

Rationale Route-tracing stereotypy is a powerful behavioral correlate of striatal function that is difficult to quantify. Measurements of route-tracing stereotypy in an automated, high throughput, easily quantified, and replicable manner would facilitate functional studies of this central nervous system region.

Objective We examined how t-pattern sequential analysis (Magnusson *Behav Res Meth Instrum Comput* 32:93–110, 2000) can be used to quantify mouse route-tracing stereotypies. This method reveals patterns by testing whether particular sequences of defined states occur within a specific time interval at a probability greater than chance.

Electronic supplementary material The online version of this article (doi:10.1007/s00213-007-0994-6) contains supplementary material, which is available to authorized users.

Stephen J. Bonasera and A. Katrin Schenk contributed equally to this work.

S. J. Bonasera (✉)
Division of Geriatrics, Department of Medicine,
University of California, San Francisco,
1550 4th St., Box 2822, San Francisco, CA 94158-2822, USA
e-mail: Steve.Bonasera@ucsf.edu

L. H. Tecott
Department of Psychiatry,
University of California, San Francisco,
San Francisco, CA 94158-2822, USA

S. J. Bonasera · A. K. Schenk
Memory and Aging Center, Department of Neurology,
University of California, San Francisco,
San Francisco, CA 94134-1207, USA

E. J. Luxenberg
Redwood High School,
Larkspur, CA 94939, USA

Results Mouse home-cage locomotor patterns were recorded after psychostimulant administration (GBR 12909, 0, 3, 10, and 30 mg/kg; *d*-amphetamine, 0, 2.5, 5, and 10 mg/kg). After treatment with GBR 12909, dose-dependent increases in the number of found patterns and overall pattern length and depth were observed. Similar findings were seen after treatment with *d*-amphetamine up to the dosage where focused stereotypies dominated behavioral response. For both psychostimulants, detected patterns displayed similar morphological features. Pattern sets containing a few frequently repeated patterns of greater length/depth accounted for a greater percentage of overall trial duration in a dose-dependant manner. This finding led to the development of a t-pattern-derived route-tracing stereotypy score. Compared to scores derived by manual observation, these t-pattern-derived route-tracing stereotypy scores yielded similar results with less within-group variability. These findings remained similar after reanalysis with removal of patterns unmatched after human scoring and after normalization of locomotor speeds at low and high ranges.

Conclusions T-pattern analysis is a versatile and robust pattern detection and quantification algorithm that complements currently available observational phenotyping methods.

Keywords Behavioral pattern detection · Sequential analysis · Behavioral pattern differentiation · T-pattern · Stereotypy, route-tracing · Behavioral phenotyping · Mouse · GBR 12909 (vanoxerine) · *d*-amphetamine

Introduction

Stereotypies are a form of abnormal repetitive behavior defined as motor actions of unknown functional purpose

that are repeatedly performed in a near-identical manner (Garner 2005; Garner and Mason 2002). Although stereotypies can be evoked by diverse environmental and pharmacological interventions (Fowler et al. 2003, 2007; Powell et al. 2000; Presti et al. 2004; Turner et al. 2003) and are associated with perturbations in multiple neurotransmitter systems (e.g., Barwick et al. 2000; Larson et al. 1996; Pogorelov et al. 2005; Presti et al. 2004; Toyota et al. 2002), stereotypical behaviors appear to reflect aberrant function within the striatum (Joyce and Iversen 1984; Saka et al. 2004; Szostak et al. 1989). Examples of these stereotypical behaviors include syntactic grooming in rodents (Aldridge and Berridge 1998; Cromwell et al. 1998) and repetitive head movements in both rats (Rebec et al. 1997) and rhesus monkeys (Saka et al. 2004).

Repeated forays along a defined path are a form of motor sequence response often observed after psychostimulant administration. Frequent repetition of these forays within a defined arena are termed route-tracing stereotypies (Cooper and Nicol 1991). However, precisely quantifying route-tracing stereotypy is difficult (e.g., Randrup and Munkvad 1967). Intensity-based stereotypy scales (e.g., Canales and Graybiel 2000; Chartoff et al. 2001; Costall et al. 1972; Creese and Iversen 1973; Ellinwood and Balster 1974, Yates et al. 2007), while commonly employed, do not quantify important aspects of route-tracing stereotypy, such as path morphology and repetitions. Stereotypy scoring and quantification are also difficult to standardize, and both intra- and inter-rater score reliabilities are problematic (Bakeman and Gottman 1997; Rebec and Bashore 1984).

Ideally, route-tracing stereotypy measures should be agnostic about pattern morphology and organization, reflect pattern temporal structure, be amenable to simple statistical tests, and be easily implemented and automated. Prior studies quantifying route-tracing stereotypies emphasized path shape and statistical properties. Route-tracing behavior, defined as repetitive sequences of motor activity occurring along well-defined routes, was first demonstrated in rats receiving amphetamine (Schjörning 1971, 1979) and first quantified using a simple four-state Markov model to produce maximum likelihood estimates of animal path repetitions within a single behavioral trial (Mueller et al. 1989). This approach, however, was not applied to locomotor patterns of arbitrary morphology and did not address pattern temporal structure, a fundamental limitation of simple Markov models. Algorithms that identify individual bouts of locomotor activity have also been reported (Drai et al. 2000; Drai and Golani 2001; Golani et al. 1999; Kafkafi et al. 2003), but these techniques are not readily adaptable to the quantification of route-tracing stereotypies. Finally, in a novel use of ergodic theory, Paulus and colleagues demonstrated that both metric and topological entropy parameters could describe increased home-cage

stereotypical locomotor behavior after psychostimulant treatment (Paulus et al. 1990, 1999; Paulus and Geyer 1991). Although these entropy measures are useful metrics for quantifying the degree of randomness present in an observed behavior, it would be useful to have assessments of the existence and characteristics of locomotor patterns that may be present.

One mathematical approach that may be particularly suitable for identifying episodes of route-tracing stereotypy in behavioral data is t-pattern sequential analysis. T-pattern analysis (for detailed exposition of technique, see Magnusson 2000) focuses on determining whether arbitrary events x_1 and x_2 in a symbolic string of $\{x_i\}$ events sequentially occur within a specified time interval at a rate greater than that expected by chance. This type of data is commonly acquired during observational behavioral studies, traditionally by manual inspection of animal activities. T-pattern analysis has been used in a wide variety of observational studies, including serum testosterone concentration in prospective fathers (Hirschenhauser et al. 2002), stereotypic decision behavior in schizophrenic patients (Lyon et al. 1994), mitral cell firing patterns in rat olfactory bulb after presentation of test odors (Nicol et al. 2005), cooperative behavior between humans and dogs when constructing an object (Kerepesi et al. 2005), and analysis of soccer team play (Borrie et al. 2002). The common feature uniting all of the above problems is the need to identify repeated behavior patterns that may irregularly occur within a period of observation. This same approach may thus be particularly well suited for identifying behavioral stereotypies.

We discuss the use of this automated pattern analysis approach to identify episodes of mouse route-tracing behavior within a home-cage environment. We demonstrate that t-pattern analysis can detect, in a sensitive manner, route-tracing stereotypies evoked by systemic treatment with psychostimulant drugs. We show that a t-pattern-derived route-tracing stereotypy score has less within-group variability compared to a manual observation-derived stereotypy score and that the patterns detected using this analytic technique are concordant with those that would be noted with visual observation. Finally, we show that the overall route-tracing stereotypy score is insensitive to the locomotor distance traveled within the test arenas.

Materials and methods

Drugs 1-[2-[bis(4-Fluorophenyl)methoxy]ethyl]-4-[3-phenylpropyl]piperazine dihydrochloride (GBR 12909; RBI/Sigma) was dissolved in normal saline vehicle with heating and vortexing. *d*-Amphetamine (RBI/Sigma) was dissolved in normal saline vehicle. Drugs were administered via intraperitoneal (i.p.) route in a 10 μ l/g injection volume. Doses

used were 3, 10, and 30 mg/kg (GBR 12909) or 2.5, 5, and 10 mg/kg (*d*-amphetamine).

Route-tracing stereotypy studies Male C57BL/6J mice (Jackson Laboratories, Bar Harbor, ME, USA), 2–3 months of age, were weighed and individually housed in standard polypropylene low-profile cages (18.5×29×13 cm) 1 week before study. Cages contained standard wood-chip mouse bedding; mice had free access to water and chow at all times except during filming. Animals were maintained on a 12-h light/dark cycle; average light intensity 550 lx. The rooms received no human traffic except during experimental testing, approximately 2 h per day. All testing was performed approximately 4–6 h before the onset of the dark cycle. A white-noise generator maintained an ambient background noise of 70 dB. Both institutional and federal regulations regarding animal care and welfare were followed (National Research Council 1996; Office of Laboratory Animal Welfare 2002, UCSF Institutional Animal Care and Use Committee).

For 6 days preceding testing, all mice received an i.p. injection of vehicle to acclimate them to handling and injection. On the day of testing, one cohort received i.p. injections of 0, 3, 10, and 30 mg/kg GBR 12909 ($n=8$ per dosage). We were unable to obtain a full data set for two mice receiving GBR 12909 3 mg/kg and one mouse receiving GBR 12909 10 mg/kg. A second cohort of mice received i.p. injections of 0, 2.5, 5, and 10 mg/kg *d*-amphetamine ($n=8$ per dosage). After injection, mice were returned to their home cage. Approximately 5 (mice treated with *d*-amphetamine) or 90 (mice treated with GBR 12909) min after injection, home cages were removed from the rack, placed on a white surface, and videotaped from 125 cm overhead. Locomotor paths were tracked at a 30 Hz sampling rate using EthoVision (Noldus, Leesburg, VA, USA), a commercially available video tracking system. Locomotor distances traveled within the testing arenas were determined using EthoVision. Individual trials from the GBR 12909-treated cohort also were observer-scored in a blinded fashion using a modified Creese–Iversen stereotypy scale (Creese and Iversen 1973) ranging from 0 (mouse either sleeping or inactive) to 6 (mouse engaged in continuous and nonstop route-tracing stereotypic behaviors). Behaviors were coded every 10 s for the total trial duration.

Data binning The first step in the automated evaluation of the stereotypic properties of locomotor paths was to divide the mouse home cage into a discrete number of regions (bins). Binning converted movement into a symbolic string of bin identity as a function of time. We calculated equal weight (maximum entropy) bins, where bin area was allowed to vary while containing similar numbers of data points per bin. To simplify calculations, we imposed a “box constraint” such that bins were set as rectangles sharing a

common row but not column [i.e., bins on the same row shared a common upper and lower boundary, but bins along the same column (with the exception of the physical arena boundaries) did not share a common boundary; Fig. 1]. Study results were similar regardless of whether a 9- (3×3) or 16-rectangle (4×4) binning scheme was used (data not shown). Raw locomotor (x,y) coordinate data from the video tracking system were imported into MATLAB (MathWorks Inc, Natick, MA, USA) for binning.

Pattern detection Once locomotor data were binned and converted to event form, we used a commercially available program (Theme, Noldus) to implement the t-pattern method for evaluating pattern structure. Individual t-patterns are detected in a hierarchical fashion. For example, assume event {b} occurs after event {a} three times. This process gives rise to three time intervals starting with {a} and ending with {b}. The algorithm calculates the maximum and minimum durations of these intervals. One then tests the null hypothesis that these durations are similar to those that would arise from a random distribution of events {a} and {b}. If the null hypothesis is rejected, {a,b} forms a pattern. Now suppose that event {c} follows the newly found pattern {a,b} more than once. The durations between the start of {a,b} and {c} are tested for significance in an identical manner. If {a,b,c} is found to be significant, it also constitutes a pattern. Events that occur after pattern {a,b,c} are tested for significance and added to the list of patterns in the same way. This process is repeated as long as events can be added to the pattern. We wrote MATLAB code that checked and validated this program’s key algorithms. As discussed in “Results,” all two-state patterns were excluded from study. A number of secondary parameters guide the algorithm during pattern identification; the specific settings for these parameters are provided in the supplemental methods (also see Supplemental Figure 1).

Pattern composition Pattern detection, as described above, detects all possible patterns present within a given trial of observational data. Each pattern consists of a set of sequentially occurring symbols (each symbol represents a bin) and a start and end time for the pattern. For instance, a detected pattern (P1) that starts at T_i , consists of the symbols {a,c,e,d}, and ends at T_f means that at T_i , the mouse entered bin ‘a’, stayed there for some time ΔT_a , then moved to bin ‘c’ for ΔT_c , then to bin ‘e’ for ΔT_e , then to bin ‘d’ for ΔT_d . At the end of ΔT_d , the mouse exits bin ‘d’ (at time T_f), and the pattern is over. Many of the patterns detected by the t-pattern algorithm will be overlapping in time. For instance, using the example pattern above, if the algorithm also finds a pattern (P2) that starts at $T_i + \Delta T_a$, consists of the symbols {c,e,d,f,g}, and ends at some time $T > T_f$, then the patterns P1 and P2 are overlapping. To determine the amount of

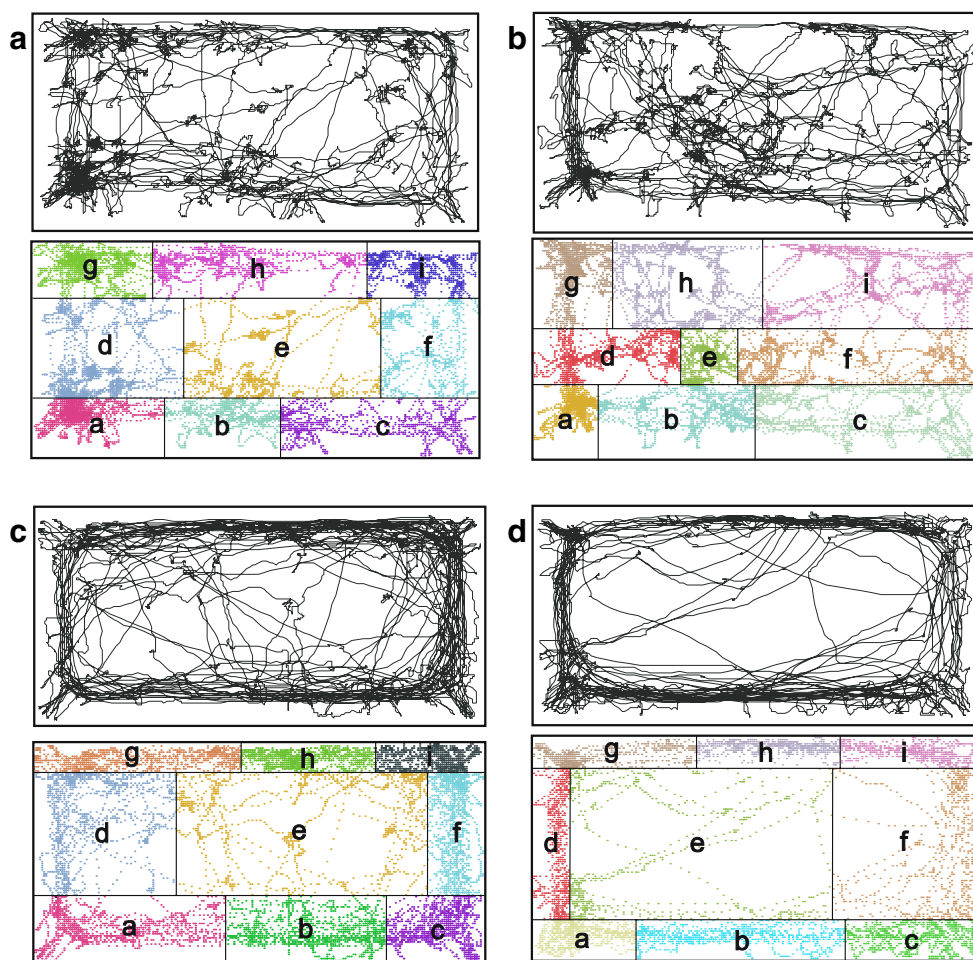
Fig. 1 Maximum entropy binning of locomotor paths using box constraints. This algorithm converts locomotor data from position as a function of time to bin as a function of time.

a Representative locomotor path and bin structure for mouse receiving vehicle injection.

b 3 mg/kg GBR 12909.

c 10 mg/kg GBR 12909.

d 30 mg/kg GBR 12909. Note that increasing doses of GBR 12909 organized locomotor behavior along the arena boundaries, a characteristic of route-tracing stereotypy



stereotyped behavior within a given trial, it is equivalent to ask what percentage of the total trial time is the animal engaging in patterned behavior. To answer this, we must use only nonoverlapping sets of patterns occurring within the trial. We will call any set of nonoverlapping, sequentially occurring patterns a pattern “composition.” For each composition, we can define the fraction of time spent in patterned behavior as the sum of all the time spent in the patterns in the composition divided by the total trial time. For each trial, there are many (depending on the number of detected patterns) possible sets of nonoverlapping patterns and thus a large number of possible compositions. To define our measure of route-tracing stereotypy, we find, for each trial, the composition that has the highest fraction of patterned behavior. This procedure gives the “maximal composition” for this trial. The maximal composition will thus be the largest group of nonoverlapping, sequentially occurring patterns that cover the most time within the trial.

Pattern composition as implemented by Theme employs a combinatorial algorithm and thus identifies maximal compositions in an unambiguous manner. For most of our observational trials, we had no difficulty rapidly obtaining a maximal pattern composition. However, for seven individ-

ual trials (six from mice receiving either 10 or 30 mg/kg GBR 12909, one from a mouse receiving *d*-amphetamine 5 mg/kg), the detection algorithm found more than 1,200 distinct patterns; all of these cases required either extended computational time to produce a maximal composition (six cases) or would not converge to a final answer. Data from this single, nonconverging trial were subsequently excluded from the following analyses. The final GBR 12909 data set thus contained a total of 28 trials: eight vehicle, six 3 mg/kg, seven 10 mg/kg, and seven 30 mg/kg. The final *d*-amphetamine data set contained 32 trials balanced across vehicle, 2.5, 5, and 10 mg/kg dosages.

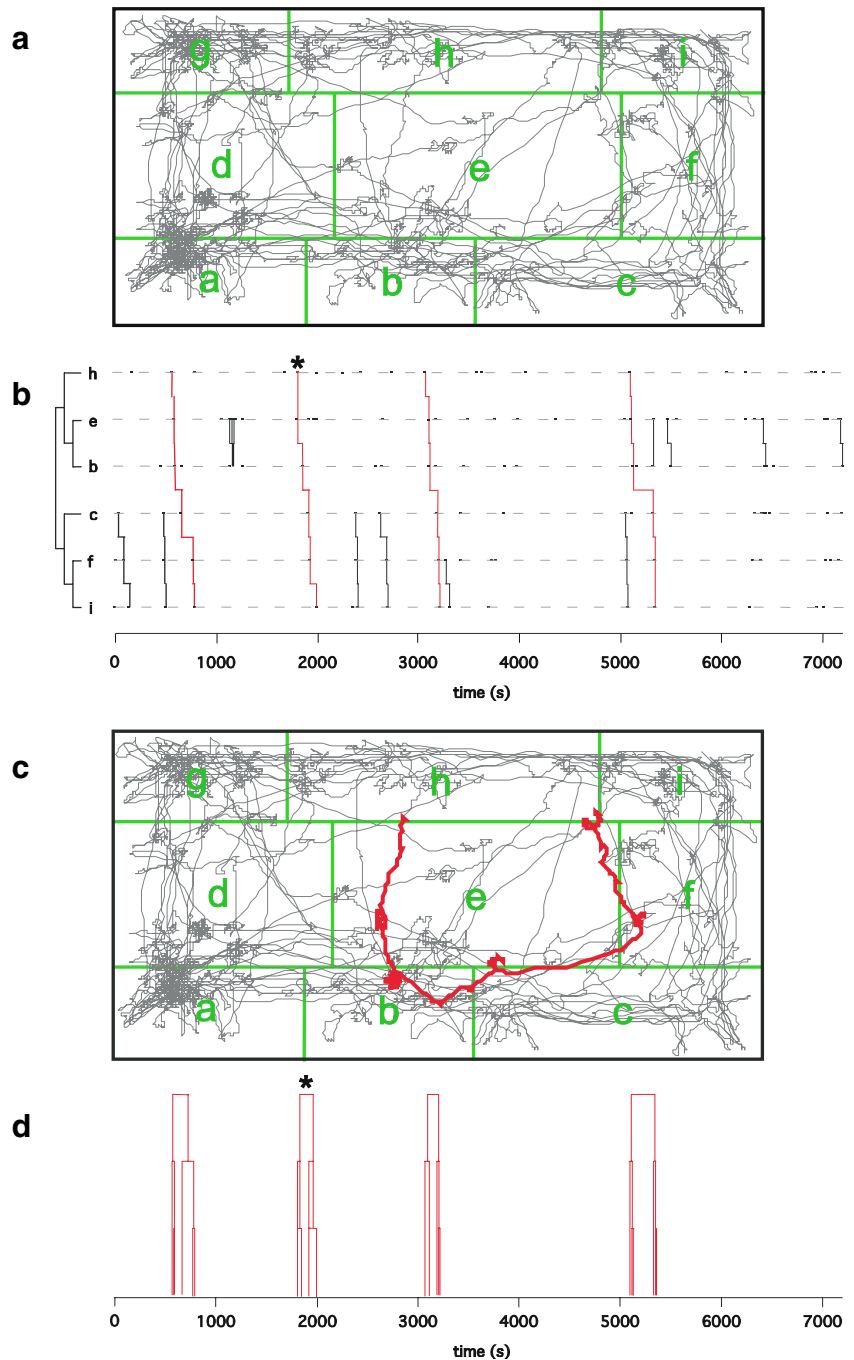
Determination of route-tracing stereotypy measures Route-tracing stereotypies can be mathematically modeled by evaluating the pattern composition for a given trial. Greater percentages of time spent in specific pattern(s) suggest greater time performing stereotyped locomotion. Thus, comparing maximal pattern compositions may provide a quantitative, easily automated means to measure route-tracing stereotypies in a variety of different conditions. A t-pattern-derived route-tracing stereotypy score was simply calculated by taking the total duration of each pattern

within the maximal composition, summing these durations, and dividing this result by the total trial duration. Thus, a behavioral trial where no patterns were identified would have a score of 0; a behavioral trial where the total duration of all patterns in the maximal composition was half the trial duration would have a score of 0.5, etc.

Pattern validation It is important that patterns identified in this automated manner have face validity when compared to patterns that may be identified by an experienced human

observer. This means that patterns identified by the automatic algorithm as the same pattern should all look similar to a human observer. Scoring sheets showing the locomotor paths traced by all patterns included in the maximal composition of an individual behavioral trial were produced for all GBR 12909 trials (see Supplemental Figure 2 for examples of instances of locomotor paths used in scoring). Start and stop points for each locomotor path were highlighted. Scoring sheets were produced for all mouse trials. Three investigators (who were blind to both

Fig. 2 Determination of individual t-patterns from locomotor paths. **a** Locomotor path (from Fig. 1a) with superimposed maximum entropy binning. **b** T-pattern testing for sequence of {h,e,b,c,f,i}. The diagram at the right depicts transitions from these states over the trial duration (transitions from the three states not included in the pattern are not shown). Note that the pattern of {h,e,b,c,f,i} is formed by combining t-patterns of {h,e,b} and {c,f,i}. Full elaborations of the pattern are depicted in red in the transition diagram; partial elaborations are depicted in black. **c** Locomotor path corresponding to second occurrence of {h,e,b,c,f,i} pattern (note asterisk). Other occurrences of this pattern trace similar paths. **d** Dendrogram plot of {h,e,b,c,f,i} pattern vs time. This plot depicts the temporal occurrence of the pattern, as well as an idealized representation of its length and depth. Note that this relatively long/deep pattern is repeated only four times throughout the trial, and accounts for a relatively small percentage of total trial duration



the mouse treatment status and how the patterns for each behavioral trial were classified within the composition) then reviewed the scoring sheets in the following manner. Locomotor paths were examined within a trial as if they were characters in a foreign alphabet, with the goal being to match the same characters (despite small differences that may be attributable to different “handwriting styles”; e.g., grouping ‘A’ with ‘A’) while avoiding false matches (e.g., not grouping a ‘u’ with a ‘v’). Patterns identified as false matches were excluded from analysis. Route-tracing stereotypy scores were then recalculated.

Compensation for locomotor differences Over the 12-min trial duration, mice receiving 10 and 30 mg/kg GBR 12909 treatment had, on average, twice as much locomotion (measured by distance) within their home cages as compared to vehicle and 3 mg/kg GBR 12909-treated mice (Supplemental Figure 3). To assess how sensitive the t-pattern derived route-tracing stereotypy scores were to locomotor distance, two different locomotor compensations were performed: one to normalize the high-dose group to low-dose locomotor distances, and the other to normalize the low-dose group to high-dose locomotor distances. Each trial from mice receiving 10 and 30 mg/kg GBR 12909 was split into two 6-min trials. Each of the resulting 6-min trials had approximately the same amount of locomotion as a complete, 12-min vehicle or 3 mg/kg GBR 12909 trial. These 6-min trials were then “re-expanded” to 12-min durations by interpolating a new position between every two position points in the trial. Route-tracing stereotypy scores were then recalculated for each of these “expanded” trials. The route-tracing stereotypy scores from the two 6-min trials that came from a single 12-min trial were then averaged to determine a score normalized for less locomotion. Similarly, two 12-min long trials from mice receiving vehicle (or 3 mg/kg GBR 12909) were “spliced” together, forming a 24-min trial. This 24-min trial thus had approximately the same amount of locomotion as the 10- or 30-mg/kg trials. These 24-min trials were then “contracted” to 12-min durations by removing every other data point. Route-tracing stereotypy scores were then recalculated for these “contracted” trials to determine a score normalized for greater locomotion.

Results

T-pattern analysis detects large scale stereotypic movements throughout the arena Representative locomotor paths within the testing arena are shown in Fig. 1a–d corresponding to GBR 12909 dosages of vehicle, 3, 10, and

30 mg/kg, respectively. The locomotor traces within Fig. 1c and d have been shortened to display the same approximate locomotor distance as observed in Fig. 1a and b. It is clear from visual inspection of these locomotor paths that mouse arena behavior is increasingly organized into circular forays around the arena perimeter with progressively increasing dosages of GBR 12909. Similar results were obtained in a second mouse cohort after *d*-amphetamine treatment.

The process of deriving an individual pattern from the locomotor pathways is depicted in Fig. 2. Figure 2a shows the locomotor trace with superimposed maximum entropy bins for one animal from the vehicle treatment group (this example comes from the same animal shown in Fig. 1a). Figure 2b depicts a schematic dendrogram for one pattern detected with the algorithm: {h,e,b,c,f,i}. The algorithm also identified 18 other three-state or longer significant patterns that are not shown. This dendrogram is a simplified way of describing the sequence of bins traversed by the mouse within a given pattern. The transition diagram of Fig. 2b further reveals that the specific pattern occurs four times. This transition diagram also shows that parts of the pattern forming the most distal regions of the dendrogram (e.g., {h,e,b} and {c,f,i}) share the strongest temporal relationships. The locomotor path corresponding to the second repetition of this identified pattern is highlighted in red in Fig. 2c (asterisk); the other repetitions of this pattern produce similar locomotor paths in the arena (data not shown). The dendrogram vs time plot for this particular pattern (Fig. 2d) is a shorthand way of depicting both pattern morphology and temporal sequence within a behavioral trial. Subsequent data will often be depicted in this format. All of these patterns were detected at $p < 0.001$. Furthermore, comparison of these detected patterns with patterns formed from the same data set after randomization shows that three state or longer/deeper patterns are not formed by chance (data not shown).

Several metrics can be defined to describe the set of detected patterns: total number of patterns, pattern length (number of states in each pattern), and pattern depth (number of levels in each pattern dendrogram). The average number of detected patterns (Supplemental Figure 4a) clearly increases with increasing doses of GBR 12909 and increases to the point where focused stereotypies dominate the behavioral response after *d*-amphetamine administration (Supplemental Figure 4c). Histograms of pattern length and depth for all trials (Supplemental Figure 4b and d for GBR 12909 and *d*-amphetamine, respectively) reveal similar dose-dependant increases in pattern length and depth. The simplest patterns are two states long (or one level deep) and uninformative; separate analysis (data not shown) demonstrates that these simple, two-state patterns often arise by chance alone. Accordingly, all further analyses will exclude any contributions from simple, two-state patterns.

Repetition of a small number of long/deep patterns best models the increased route-tracing stereotypy evoked by GBR 12909 and d-amphetamine The above results demonstrate a dose-dependant increase in the total number of detected patterns and pattern length/depth with increasing psychostimulant dosage. However, a more useful measure of route-tracing behavior would involve choosing from the total group of identified patterns a subset of nonoverlapping patterns that maximized the percentage of trial duration spent within patterned locomotion. In other words, the behavioral time line would be “tiled” with the largest set of nonoverlapping patterns. One would thus predict that in a trial characterized by little route-tracing locomotion (such as after vehicle injection), many short, shallow patterns will be identified and, when placed in a nonoverlapping manner across the trial duration, account for a relatively small percentage of overall trial duration. By contrast, one would predict that in a trial characterized by frequent route-tracing

stereotypies (such as after psychostimulant injection), time spent in a small number of long/deep but frequently repeating (and nonoverlapping) patterns will account for a significant percentage of trial duration.

Figure 3a–d depicts the best (maximal composition) sets of nonoverlapping patterns for representative mice receiving vehicle, 3, 10, and 30 mg/kg GBR 12909 treatment (depicted locomotor paths within this figure are cartoons of the identified patterns). As this figure clearly demonstrates, patterns cover greater periods of the overall observation with increased dosages of GBR 12909. Furthermore, this figure also suggests that the increased pattern coverage is accompanied by a decrease in the total number of patterns within the maximal composition. Similar findings are observed after *d*-amphetamine treatment. At the highest psychostimulant doses, mouse paths frequently repeated a small number of relatively long patterns. These mice spent more time performing a smaller range of observed behaviors.

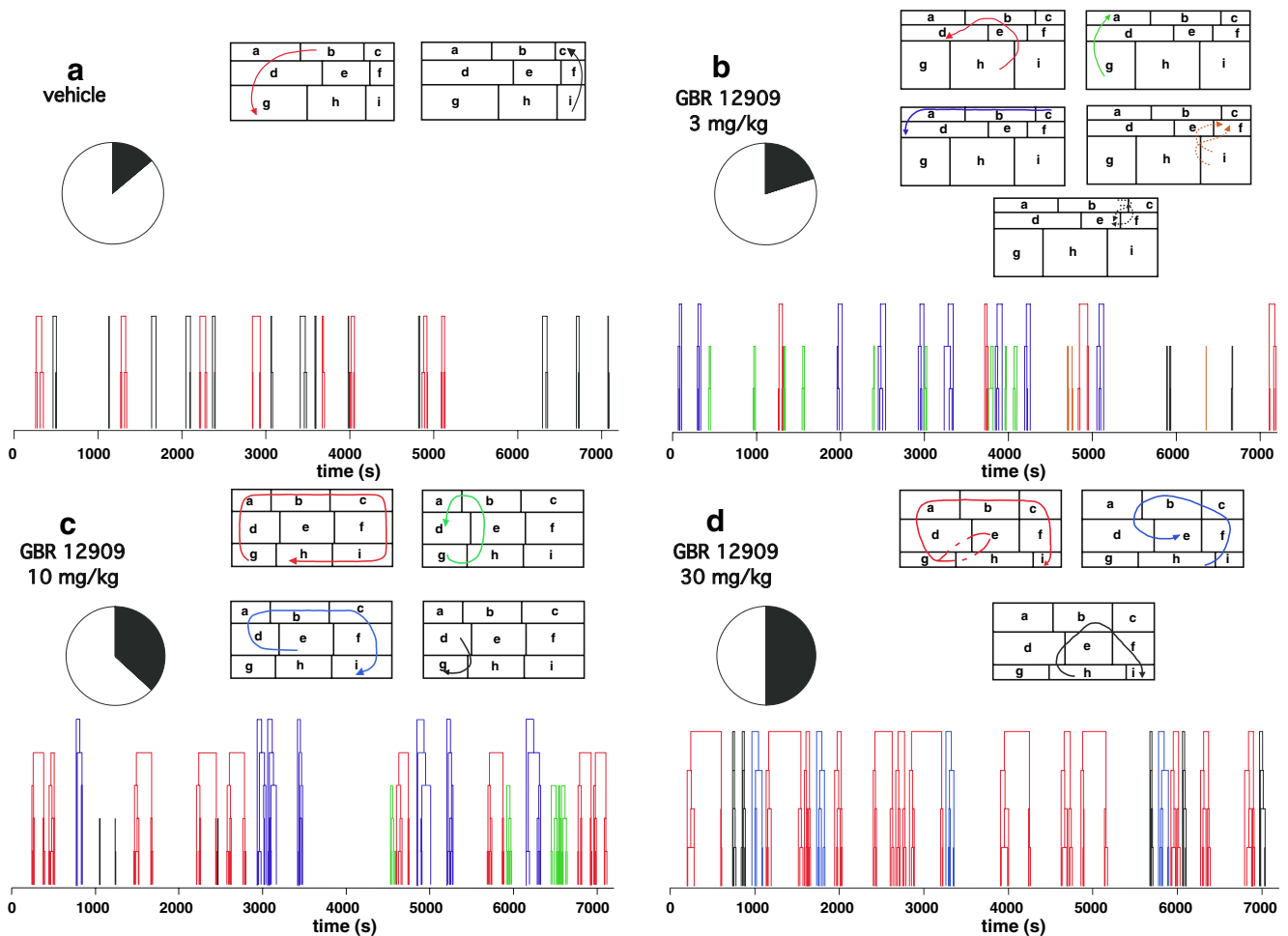


Fig. 3 Pattern composition: selecting the subset of nonoverlapping, detected patterns to cover the greatest fraction of trial duration. Representative examples for vehicle (a), GBR 12909 3 mg/kg (b), 10 mg/kg (c), 30 mg/kg (d). Note trend of increasing pattern coverage of overall trial using fewer patterns as GBR 12909 dosage increased. Drawings depict each pattern included in the maximal composition;

color of the patterns in these drawings corresponds to dendrogram color in the dendrogram vs time plot (time in seconds). Dotted lines represent patterns where a state within the dendrogram was equally likely to transition into one of two immediately adjacent bins. Pie chart depicts fraction of total trial duration accounted for by patterns

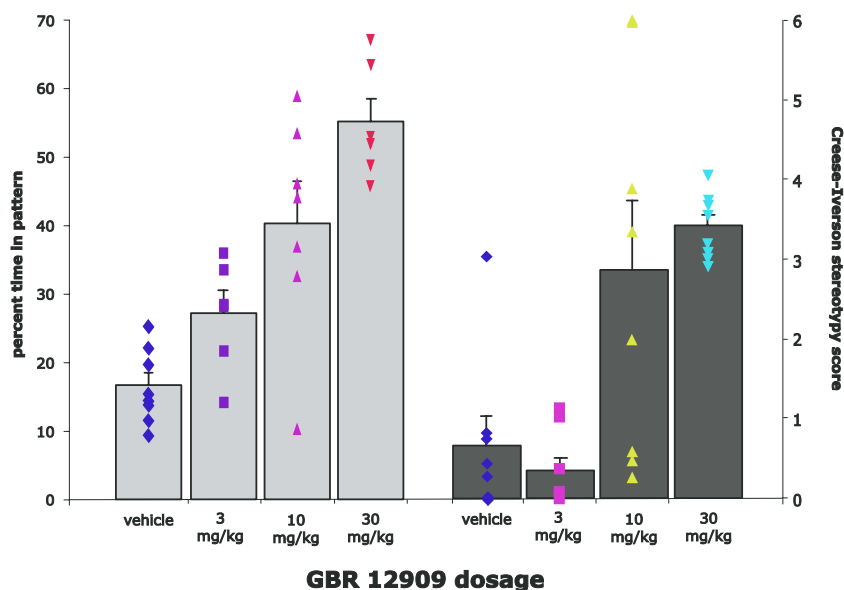


Fig. 4 Comparison of t-pattern-derived route-tracing stereotypy score (light grey bars) with manual observation-derived stereotypy score (dark grey bars). T-pattern-derived route-tracing stereotypy scores increase with greater doses of GBR 12909. Individual route-tracing stereotypy score values for each dosage overlaid on respective bar. Error bars are ± 1 standard error. Note significantly less within-group variability when analysis performed using t-patterns. One-way

ANOVA on dosage effect for t-pattern derived route-tracing stereotypy score $F_{3,24}=48.53$ ($p<0.00001$, $r^2=0.78$, all pairwise comparisons significant by Duncan's multiple range test); for human observer-derived (Creese–Iverson) stereotypy score $F_{3,27}=9.65$ ($p<0.0002$, $r^2=0.21$, pairwise comparisons between vehicle and 10 mg/kg, vehicle and 30 mg/kg, 3 and 10 mg/kg, 3 and 30 mg/kg groups significant by Duncan's multiple range test)

Thus, higher route-tracing stereotypy scores reflect a greater predictability of mouse behavior. High route-tracing stereotypy scores also reflect a psychostimulant-induced simplification of locomotor behavior.

Figures 4 and 5 demonstrate that these findings extend to the analysis of both tested cohorts. One-way analysis of variance (ANOVA) demonstrated a highly significant effect of increased GBR 12909 dosage on pattern coverage of the observational trial timeline (Fig. 4, light grey bars). One-way ANOVA also shows that the locomotor behavior observed after increasing doses of GBR 12909 is characterized by smaller pattern sets (Supplemental Figure 5). Furthermore, results from the pattern composition analysis were concordant with results obtained by manual scoring of the video data (Fig. 4, dark grey bars). Note that while the route-tracing stereotypy trends determined by both t-pattern analysis and manual observation both suggest increasing route-tracing behavior with increasing GBR 12909 dosage, within-group variability of the data when analyzed by t-pattern methods is far less than within-group variability when analyzed by manual observation. This suggests that even an experienced observer is able to score these locomotor data with less precision compared to the t-pattern approach. Finally, one-way ANOVA demonstrated a highly significant effect of increased *d*-amphetamine dosage on pattern coverage of the observational trial timeline (Fig. 5). T-pattern-derived stereotypy scores thus reflect the increasing route-tracing stereotypy observed after treatment with two common psychostimulants: GBR 12909 and *d*-amphetamine.

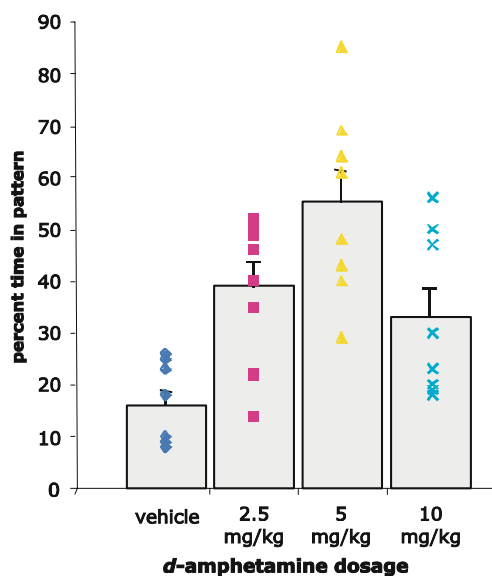


Fig. 5 T-pattern-derived route-tracing stereotypy scores increase with greater doses of *d*-amphetamine. Individual route-tracing stereotypy score values for each dosage overlaid on respective bar. Error bars are ± 1 standard error. One-way ANOVA on dosage effect for t-pattern derived route-tracing stereotypy score $F_{3,28}=9.91$ ($p<0.0001$, $r^2=0.51$, pairwise comparisons between vehicle and 2.5 mg/kg and vehicle and 5 mg/kg, significant by Duncan's multiple range test). Decreased route-tracing stereotypy score observed in group receiving *d*-amphetamine 10 mg/kg reflects the development of the amphetamine response stationary phase (Schjörring 1971; also referred to as focused stereotypy, Canales and Graybiel 2000) as the trial progressed

Locomotor patterns identified by t-pattern analysis have face validity It is reasonable to question whether the locomotor patterns identified above actually correspond to what a trained observer might characterize as a stereotypic pattern. Supplemental Figure 2 depicts representative pattern scoring (as described in methods) for trials involving a mouse treated with vehicle (a) or GBR 12909 30 mg/kg (b). Supplemental Figure 6 depicts analogous representative patterns involving a mouse treated with vehicle (a) or *d*-amphetamine 10 mg/kg (b). As these figures demonstrate, sets of locomotor paths representing different instances of the same identified pattern show similar overall morphologies. Paths “crossed off” with a grey ‘x’ designate patterns whose morphologies did not match other paths in the pattern group. These were determined by consensus of three raters. The process of rejecting these patterns had approximately equal effects across all doses, removing 30 of 108 patterns in the vehicle group (28%), 49 of 137 patterns in the GBR 12909 3 mg/kg group (36%), 51 of 187 patterns in the GBR 12909 10 mg/kg group (27%), 46 of 260 patterns in the GBR 12909 30 mg/kg group (18%), and 176 of 692 patterns overall (25%). More importantly, the process of rejecting unconfirmed patterns had a relatively minor effect on the overall analysis (Supplemental Figure 2d). While this process decreased the time spent in stereotypic behavior between 30 and 50%, this decrease was evenly distributed across all treatment groups. Of note, after pattern rejection, the difference in route-tracing stereotypy scores between vehicle and 3 mg/kg treatment groups lost statistical significance; however, these two groups were not found to be statistically different by manual scoring (Fig. 4).

Route-tracing stereotypy scores are not correlated with overall locomotor activity An important issue that would limit the utility of the preceding t-pattern approach would be if increased locomotor activity itself (and the concomitant increase in overall detected patterns) led to a larger route-tracing stereotypy score. Investigating this is particularly challenging, as there are no pharmacological treatments known to fully dissociate changes in overall locomotor speed from the development of route-tracing stereotypies (e.g., Griebel et al. 2000; Kliethermes and Crabbe 2006; Ohl et al. 2001). As per the rationale described in “Materials and methods,” we chose to mathematically transform home-cage data to normalize locomotor speeds without altering mouse locomotor paths. The results of these analyses are shown in Supplemental Figure 7. No significant effect of greater locomotor distances on route-tracing stereotypy scores was found. Thus, t-pattern-derived route-tracing stereotypy scores are insensitive to overall mouse locomotion within the home cage over a broad range of potential total locomotor distances.

Discussion

Our findings suggest that t-pattern sequential analysis quantifies route-tracing stereotypies in a reliable, replicable, and automated manner. This mathematical approach detected increases in overall locomotor pattern length/depth and total number of patterns in mice treated with the psychostimulants GBR 12909 and *d*-amphetamine, drugs both known to evoke route-tracing stereotypies. Furthermore, this approach also detected dose-dependent increases in mouse route-tracing stereotypy after treatment with both drugs. Notably, the increased route-tracing stereotypy evoked by larger psychostimulant doses was described by the frequent occurrence of a few long/deep locomotor patterns that accounted for a significant fraction of the overall observed locomotor behavior. By contrast, mouse activity after treatment with saline vehicle was better described by the infrequent occurrence of a greater number of short and shallow locomotor patterns accounting for a smaller fraction of the overall observed locomotor behavior. At high doses of both GBR 12909 and *d*-amphetamine, the organization of locomotor behavior into relatively few, frequently repeated patterns mirrors the loss of behavioral complexity seen after psychostimulant administration. T-pattern-derived route-tracing stereotypy scores showed less variance than manual observation-derived scores at all doses. T-pattern analysis also identified locomotor patterns qualitatively similar to what a human observer might identify in a manner insensitive to the overall locomotor distance traveled during the trials.

Precise quantification of route-tracing stereotypy allows us to better visualize subtle differences in the locomotor paths evoked by GBR 12909 and *d*-amphetamine. For example, the increase in route-tracing stereotypy score evoked by *d*-amphetamine usually results from the frequent repetition of a single detected pattern. By contrast, GBR 12909-evoked increases in route-tracing stereotypy score usually result from repetition of two to four detected patterns. In all cases, our findings remain consistent with the Lyon–Robbins hypothesis regarding the behavioral effects of psychostimulants. Treatment with either agent leads to an increase in total patterns initiated, as well as a concomitant decrease in the variety of patterns observed.

Descriptions of route-tracing stereotypy ideally would quantify several features of this behavior, such as the identification of the locomotor paths repetitively traveled (i.e., the behavioral unit of repetition) and path temporal structure. In this study, the set of patterns chosen to partition each individual behavioral trial is the set of stereotypic locomotor paths taken by the mouse. Other than limits imposed by the chosen binning strategy, this approach detects paths of any morphology in a hypothesis-independent manner. Thus, this method is not limited to specific, predetermined,

and potentially arbitrary behavioral sequences. For any trial, it is straightforward to determine how many times a specific pattern occurred and what percentage of overall trial duration is contributed by each pattern. Thus, t-pattern analysis appropriately quantifies important aspects of route-tracing stereotypy. Of note, while this approach fully quantifies the locomotor patterns for individual mice during bouts of route-tracing stereotypy, we produce no measures of path similarity across/within different treatment classes. One way of addressing this issue would be to compare path topologies (which are independent of binning) using appropriately adapted machine learning/computer vision algorithms designed for character recognition.

While this analytical approach is a promising method of automating the measurement of route-tracing stereotypies, several issues influence the interpretation of these results and future implementations. As mentioned earlier, the t-pattern algorithm uses the null hypothesis that temporal events are distributed under a Poisson distribution. Statistics calculated from this distribution imply that event probabilities remain constant throughout time. For short behavioral observations (such as those employed in this study), this condition is likely true. Thus, our study did not reject the null hypothesis simply due to time variation within the data. However, event transition probabilities have been demonstrated to change as a function of circadian time (Poirel and Larouche 1989; Richardson et al. 1985). If event transition probabilities are not stationary with time, then one needs some method of windowing the data into stationary epochs that can be analyzed individually. Many methods for choosing these stationary epochs exist (Hofer et al. 2007). This process would permit one to calculate the time course of the route-tracing stereotypy score.

Another factor that may influence pattern detection is the binning of the test arena. In our analysis, the manner in which we divide the test arena leads to the emergence of “prohibited transitions” where it may be impossible to directly move from one state to another. The effect of prohibited transitions (overdispersion) on the overall result varies as the mouse moves through the arena, being at a minimum when the mouse is in the arena center and most prominent when the mouse is in one of the arena corners. The effect of overdispersion is that certain transitions may be overrepresented simply as an artifact of the binning resolution. Overdispersion may thus result in false pattern detection. To compensate for this issue, we set the algorithm to only identify highly significant patterns (occurring at p of 0.001 or less).

Before this approach can be ported to large-scale applications, a few computational issues must be addressed. The algorithm for pattern detection is computationally robust and fast and can be rapidly applied to a large data set. However, the currently implemented algorithm for

pattern composition is inefficient due to its combinatorial underpinnings. Methodologies such as genetic algorithms, evolutionary programming, and evolutionary strategies have been successfully applied to highly complex problems (e.g., Bäck 1996 for a broad review of the topic) and may prove useful in developing a more computationally efficient algorithm for pattern composition. Obtaining pattern compositions through means other than combinatorial “brute force” also requires addressing the nontrivial (and beyond the scope of this paper) problem of proving that the final composition truly is optimal.

In conclusion, we describe a novel method of employing t-pattern sequential analysis to accurately quantify the extent to which mouse locomotor behavior within an arena is expressed in repetitive patterns. The advantages of this algorithm include its ability to detect repeating patterns present within the data without a priori assumptions about specific pattern features, appropriate face validity of algorithm-identified locomotor patterns, and insensitivity of pattern identification and composition to total locomotor activity levels. Algorithm limitations that should be addressed in future studies include further refinement of critical interval calculation to account for prohibited state transitions, improved and automated identification and clustering of pattern morphologies, the addition of windowing capabilities to deal with violations of data stationarity, and improved implementation of composition algorithms’ ability to analyze larger and longer data sets. With these sorts of improvements, this approach may serve as a widely applied tool for providing rapid and reliable analyses of mouse behavior.

Acknowledgements The authors would like to thank Daniel Healy for computer and CVS support, Albert Willemsen and Wilant Van Giessen for their technical advice, Scott Carra and Kevin Yee for their technical assistance, Magnus Magnusson, Ph.D. for fruitful discussions regarding pattern detection algorithms, Charles McCulloch, Ph. D. for discussions regarding biostatistical issues, and Noldus for their generous Theme trial license. This study was supported by grants from the Stephen D. Bechtel Jr. Foundation (SJB, AKS), the Brookdale National Fellowship Program (SJB), NIA-AG026043 (SJB, AKS), NIMH-MH065983 (SJB), and NIMH-MH077128 (LHT).

References

- Aldridge JW, Berridge KC (1998) Coding of serial order by neostriatal neurons: a “natural action” approach to movement sequence. *J Neurosci* 18:2777–2787
- Bäck T (1996) Evolutionary algorithms in theory and practice—evolution strategies, evolutionary programming, genetic algorithms. Oxford University Press, Oxford, UK
- Bakeman R, Gottman JM (1997) Observing interaction. An introduction to sequential analysis. Cambridge University Press, Cambridge UK
- Barwick VS, Jones DH, Richter JT, Hicks PB, Young KA (2000) Subthalamic nucleus microinjections of 5-HT₂ receptor antagonists suppress stereotypy in rats. *Neuroreport* 11:267–270

- Borrie A, Jonsson GK, Magnusson MS (2002) Temporal pattern analysis and its applicability in sport: an explanation and exemplar data. *J Sports Sci* 20:845–852
- Canales JJ, Graybiel AM (2000) A measure of striatal function predicts motor stereotypy. *Nat Neurosci* 3:377–383
- Chartoff EH, Marck BT, Matsumoto AM, Dorsa DM, Palmiter RD (2001) Induction of stereotypy in dopamine-deficient mice requires striatal D1 receptor activation. *Proc Natl Acad Sci U S A* 98:10451–10456
- Cooper JJ, Nicol CJ (1991) Stereotypic behaviour affects environmental preference in bank voles, *Clethrionomys glareolus*. *Animal Behav* 41:971–977
- Costall B, Naylor RJ, Olley JE (1972) The substantia nigra and stereotyped behaviour. *Eur J Pharmacol* 18:95–106
- Creese I, Iversen SD (1973) Blockage of amphetamine induced motor stimulation and stereotypy in the adult rat following neonatal treatment with 6-hydroxydopamine. *Brain Res* 55:369–382
- Cromwell HC, Berridge KC, Drago J, Levine MS (1998) Action sequencing is impaired in D_{1A}-deficient mutant mice. *Eur J Neurosci* 10:266–2432
- Drai D, Golani I (2001) SEE: a tool for the visualization and analysis of rodent exploratory behavior. *Neurosci Biobehav Rev* 25:409–426
- Drai D, Benjamini Y, Golani I (2000) Statistical discrimination of natural modes of motion in rat exploratory behavior. *J Neurosci Methods* 96:119–131
- Ellinwood EH Jr, Balster RL (1974) Rating the behavioral effects of amphetamine. *Eur J Pharmacol* 28:35–41
- Fowler SC, Birkestrand B, Chen R, Vorontsova E, Zarcone T (2003) Behavioral sensitization to amphetamine in rats: changes in the rhythm of head movements during focused stereotypies. *Psychopharmacology (Berl)* 170:167–177
- Fowler SC, Covington HE 3rd, Miczek KA (2007) Stereotyped and complex motor routines expressed during cocaine self-administration: results from a 24-h binge of unlimited cocaine access in rats. *Psychopharmacology (Berl)* 192:465–478
- Garner JP (2005) Stereotypies and other abnormal repetitive behaviors: potential impact on validity, reliability, and replicability of scientific outcomes. *ILAR J* 46:106–117
- Garner JP, Mason GJ (2002) Evidence for a relationship between cage stereotypies and behavioural disinhibition in laboratory rodents. *Behav Brain Res* 136:83–92
- Golani I, Kafkafi N, Drai D (1999) Phenotyping stereotypic behaviour: collective variables, range of variation and predictability. *Appl Anim Behav Sci* 65:191–220
- Griebel G, Belzung C, Perrault G, Sanger DJ (2000) Differences in anxiety-related behaviours and in sensitivity to diazepam in inbred and outbred strains of mice. *Psychopharmacology (Berl)* 148:164–170
- Hirschenhauser K, Frigerio D, Grammer K, Magnusson MS (2002) Monthly patterns of testosterone and behavior in prospective fathers. *Horm Behav* 42:172–181
- Hofer H, Staude G, Wolf W (2007) A method for locating gradual changes in time series. *Biomed Tech (Berl)* 52:137–142
- Joyce EM, Iversen SD (1984) Dissociable effects of 6-OHDA-induced lesions of neostriatum on anorexia, locomotor activity and stereotypy: the role of behavioural competition. *Psychopharmacology (Berl)* 83:363–366
- Kafkafi N, Lipkind D, Benjamini Y, Mayo CL, Elmer GI, Golani I (2003) SEE locomotor behavior test discriminates C57BL/6J and DBA/2J mouse inbred strains across laboratories and protocol conditions. *Behav Neurosci* 117:464–477
- Kerepesi A, Jonsson GK, Miklosi A, Topal J, Csanyi V, Magnusson MS (2005) Detection of temporal patterns in dog–human interaction. *Behav Processes* 70:69–79
- Kliethermes CL, Crabbe JC (2006) Genetic independence of mouse measures of some aspects of novelty seeking. *Proc Natl Acad Sci U S A* 103:5018–5023
- Larson J, Quach CN, LeDuc BQ, Nguyen A, Rogers GA, Lynch G (1996) Effects of an AMPA receptor modulator on methamphetamine-induced hyperactivity in rats. *Brain Res* 738:353–356
- Lyon M, Lyon N, Magnusson MS (1994) The importance of temporal structure in analyzing schizophrenic behavior: some theoretical and diagnostic implications. *Schizophr Res* 13:45–56
- Magnusson MS (2000) Discovering hidden time patterns in behavior: t-patterns and their detection. *Behav Res Meth Instrum Comput* 32:93–110
- Magnusson MS (2004) Theme: powerful tool for detection and analysis of hidden patterns in behavior. Reference Manual, Version 5.0. Noldus Information Technology, Noldus Information Technology
- Mueller K, Hollingsworth EM, Cross DR (1989) Another look at amphetamine-induced stereotyped locomotor activity in rats using a new statistic to measure locomotor stereotypy. *Psychopharmacology (Berl)* 97:74–79
- National Research Council (1996) Guide for the care and use of laboratory animals. National Academy, Washington, DC
- Nicol AU, Kendrick KM, Magnusson MS (2005) Communication within a neural network. In: Anolli L, Duncan S Jr, Magnusson MS, Riva G (eds) The hidden structure of interaction; from neurons to culture patterns. IOS Press, Amsterdam
- Office of Laboratory Animal Welfare, National Institutes of Health (2002) Public health service policy on humane care and use of laboratory animals. Department of Health and Human Services, Washington, DC
- Ohl F, Sillaber I, Binder E, Keck ME, Holsboer F (2001) Differential analysis of behavior and diazepam-induced alterations in C57BL/6N and BALB/c mice using the modified hole board test. *J Psychiatr Res* 35:147–154
- Paulus MP, Geyer MA (1991) A temporal and spatial scaling hypothesis for the behavioral effects of psychostimulants. *Psychopharmacology (Berl)* 104:6–16
- Paulus MP, Geyer MA, Gold LH, Mandell AJ (1990) Application of entropy measures derived from the ergodic theory of dynamical systems to rat locomotor behavior. *Proc Natl Acad Sci U S A* 87:723–727
- Paulus MP, Dulawa SC, Ralph RJ, Mark AG (1999) Behavioral organization is independent of locomotor activity in 129 and C57 mouse strains. *Brain Res* 835:27–36
- Pogorelov VM, Rodriguiz RM, Insko ML, Caron MG, Wetsel WC (2005) Novelty seeking and stereotypic activation of behavior in mice with disruption of the *Dat1* gene. *Neuropsychopharmacology* 30:1818–1831
- Poirel C, Larouche B (1989) Circadian patterns of basic emotional reactivity and stress related events revisited in mice treated with lithium: behavioral rhythmometric analyses. *Chronobiologia* 16:229–239
- Powell SB, Newman HA, McDonald TA, Bugenhagen P, Lewis MH (2000) Development of spontaneous stereotyped behavior in deer mice: effects of early and late exposure to complex environment. *Dev Psychobiol* 37:100–108
- Presti MF, Gibney BC, Lewis MH (2004) Effects of intrastriatal administration of selective dopaminergic ligands on spontaneous stereotypy in mice. *Physiol Behav* 80:433–439
- Randrup A, Munkvad I (1967) Stereotyped activities produced by amphetamine in several animal species and man. *Psychopharmacologia* 11:300–310
- Rebec GV, Bashore TR (1984) Critical issues in assessing the behavioral effects of amphetamine. *Neurosci Biobehav Rev* 8:153–159
- Rebec GV, White IM, Puotz JK (1997) Responses of neurons in dorsal striatum during amphetamine-induced focused stereotypy. *Psychopharmacology (Berl)* 130:343–351
- Richardson GS, Moore-Ede MC, Czeisler CA, Dement WC (1985) Circadian rhythms of sleep and wakefulness in mice: analysis using long-term automated recording of sleep. *Am J Physiol* 248:R320–R330

- Saka E, Goodrich C, Harlan P, Madras BK, Graybiel AM (2004) Repetitive behaviors in monkeys are linked to specific striatal activation patterns. *J Neurosci* 24:7557–7565
- Schiørring E (1971) Amphetamine induced selective stimulation of certain behaviour items with concurrent inhibition of others in an open-field test with rats. *Behaviour* 39:1–17
- Schiørring E (1979) An open field study of stereotyped locomotor activity in amphetamine-treated rats. *Psychopharmacology (Berl)* 66:281–287
- Szostak C, Jakubovic A, Phillips AG, Fibiger HC (1989) Neurochemical correlates of conditioned circling within localized regions of the striatum. *Exp Brain Res* 75:430–440
- Toyota H, Dugovic C, Koehl M, Laposky AD, Weber C, Ngo K, Wu Y, Lee DH, Yanai K, Sakurai E, Watanabe T, Liu C, Chen J, Barbier AJ, Turek FW, Fung-Leung WP, Lovenberg TW (2002) Behavioral characterization of mice lacking histamine H(3) receptors. *Mol Pharmacol* 62:389–397
- Turner CA, Lewis MH, King MA (2003) Environmental enrichment: effects on stereotyped behavior & dendritic morphology. *Dev Psychobiol* 43:20–27
- Yates JW, Meij JTA, Sullivan JR, Richtand NM, Yu L. (2007) Bimodal effect of amphetamine on motor behaviors in C57BL/6 mice. *Neurosci Lett* 427:66–70