

A Data Mining Approach to *In Vivo* Classification of Psychopharmacological Drugs

Neri Kafkafi^{*}1, Daniel Yekutieli² and Greg I Elmer¹

¹Department of Psychiatry, Maryland Psychiatric Research Center, University of Maryland School of Medicine, Baltimore, MD, USA; ²Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel

Data mining is a powerful bioinformatics strategy that has been successfully applied *in vitro* to screen for gene-expression profiles predicting toxicological or carcinogenic response ('class predictors'). In this report we used a data mining algorithm named Pattern Array (PA) *in vivo* to analyze mouse open-field behavior and characterize the psychopharmacological effects of three drug classes—psychomotor stimulant, opioid, and psychotomimetic. PA represents rodent movement with ~100 000 complex patterns, defined as multiple combinations of several ethologically relevant variables, and mines them for those that maximize any effect of interest, such as the difference between drug classes. We show that PA can discover behavioral predictors of all three drug classes, thus developing a reliable drug-classification scheme in small group sizes. The discovered predictors showed orderly dose dependency despite being explicitly mined only for class differences, with the high doses scoring 4–10 standard deviations from the vehicle group. Furthermore, these predictors correctly classified in a dose-dependent manner four 'unknown' drugs (ie that were not used in the training process), and scored a mixture of a psychomotor stimulant and an opioid as being intermediate between these two classes. The isolated behaviors were highly heritable ($h^2 > 50\%$) and replicable as determined in 10 inbred strains across three laboratories. PA can in principle be applied for mining behaviors predicting additional properties, such as within-class differences between drugs and within-drug dose–response, all of which can be measured automatically in a single session per animal in an open-field arena, suggesting a high potential as a tool in psychotherapeutic drug discovery.

Neuropsychopharmacology (2009) **34**, 607–623; doi:10.1038/npp.2008.103; published online 20 August 2008

Keywords: drug discovery; animal model; pattern array; SEE; mouse; locomotor behavior

INTRODUCTION

Behavioral animal models used in psychopharmacology drug discovery are often times developed strictly for their predictive validity. The behavioral endpoint may or may not be developed with specific regard to the model's face or construct validity. The main purpose of this type of model is typically to predict the neuropharmacological properties of novel compounds with reasonable degree of sensitivity and specificity (Willner, 1991). Despite the lack of direct face or construct validity, this type of approach has proved valuable in terms of its contribution to assessing the potential pharmacological properties of novel compounds (Creese *et al*, 1976; Kuribara and Tadokoro, 1981; Arnt, 1982; Yen-Koo *et al*, 1985; Wadenberg and Hicks, 1999). A drawback of many of these animal models, however, is that they are severely restricted to the identification of a narrow

pharmacological class, often times to a specific molecular mechanism. The focus on specific mechanistic interventions might prove unsatisfactory in psychiatric drug discovery, as psychiatric illness is not likely constrained to a single biological entity (Butcher *et al*, 2004; Hood and Perlmutter, 2004; Agid *et al*, 2007). Moreover, such models are not high throughput for initial *in vivo* screening of novel compounds, because classifying a compound into one of n psychopharmacological classes generally requires evaluating it with a battery of n class-specific behavioral assays (eg locomotor activity for identifying psychomotor stimulants, tail-flick for identifying opioids, etc). An *in vivo* psychopharmacological screening paradigm capable of predicting a wide range of psychopharmacological classes with sensitivity and specificity, especially using a single assay, would provide a valuable tool in the armamentarium of drug discovery (van der Greef and McBurney, 2005). A predictive high-throughput behavioral screen could identify novel chemical entities with increased efficacy and improved therapeutic profile.

When searching for predictive models *in vitro*, data mining approaches that screen multiple potential endpoints in parallel are commonly used to improve sensitivity, specificity, and generality. For example, large gene-expression

*Correspondence: Dr N Kafkafi, Department of Psychiatry, Maryland Psychiatric Research Center, University of Maryland School of Medicine, Maple and Locust Streets, Room no. A7, PO Box 21247, Baltimore, MD 21228, USA, Tel: +1 410 402 6095, Fax: +1 410 402 6066, E-mail: nkafkafi@mprc.umaryland.edu

Received 25 March 2008; revised 29 May 2008; accepted 2 June 2008

data sets are mined for expression profiles that are predictive of a toxicological or carcinogenic response ('class predictors'; Golub *et al*, 1999; Clarke *et al*, 2004; Thomas *et al*, 2007). In this instance specific mechanisms do not need to be invoked, the only criterion is that the gene(s) are highly predictive of the outcome. This type of data mining approach was recently utilized *in vitro* to discover gene-expression patterns that could be used for the classification of psychoactive drugs (Gunther *et al*, 2003, 2005). It has been proposed that the same approach would also be advantageous for *in vivo* analysis of behavioral data (Brunner *et al*, 2002; Tecott and Nestler, 2004), but no such results have been published to date.

Ideally, an animal model designed for *in vivo* psychopharmacological class prediction would be amenable to relatively high throughput and focus on behaviors that have the following properties: (1) algorithmically defined and automatically measured; (2) common enough in natural behavior to supply large samples; (3) complex enough to provide a relatively detailed profile of a drug's psychoactive properties, and (4) resistant as possible to interactions with confounding environmental changes. A behavioral platform that meets these criteria is locomotor behavior in a novel arena. The path of a mouse or a rat can be automatically measured with high precision using commercially available tracking systems. The test is technically simple to conduct and does not require previous conditioning of the animal. A single 1-hour session tracked at a rate of 30 Hz generates a path consisting of many thousands of coordinates—an information-rich data source containing a repertoire of structured movement patterns amenable to mathematical description (Drai and Golani, 2001). Certain aspects of these behaviors are highly heritable and replicable across laboratories (Kafkafi *et al*, 2005). The purpose of this report is to describe a novel application of a data mining algorithm named Pattern Array (PA; see Kafkafi *et al*, in press) to analyze results from this behavioral assay.

The main barrier to realizing a data mining approach in a behavioral assay is that the natural units of behavior are not as well understood and defined as the natural units employed by *in vitro* assays (eg highly annotated genes in gene-expression profiling). Thus the key to applying a data mining approach in behavioral analysis is designing a useful 'chip', that is, constructing a proper categorization of the data into multiple types of behavior that can be mined. To this end the units employed by PA consist of a large number (~100 000) of complex movement patterns, algorithmically defined as simultaneous combinations of several ethologically relevant 'attributes' such as the distance from the arena wall (Broadhurst, 1975; Leppanen *et al*, 2005; Lipkind *et al*, 2004), the speed and acceleration of movement (Kafkafi *et al*, 2003b), direction of movement (Horev *et al*, 2007), and turning (Kafkafi and Elmer, 2005). For example, by using the three attributes of acceleration, distance from the wall and turning, a specific movement pattern in PA may be defined as the combination 'heavily braking while moving near the wall but turning slightly away from it'. PA then considers how frequently each such pattern is used by the animals, and mines the ~100 000 patterns for those that best predict any factor of interest. For example, the specific pattern above was actually found by PA to differentiate *SOD1* mutant rats, an animal model of Amyotrophic Lateral Sclerosis, from the wild-type controls at half the age of the known disease onset (Kafkafi *et al*, in press). In the current application we utilized PA in the mouse to mine for behavioral patterns that are predictive of three major drug classes: psychomotor stimulants, opioids and psychotomimetics.

This study thus follows the typical steps of constructing a classification model in data mining (for comprehensive introduction see Tan *et al*, 2006): (1) The data, in our case consisting of path coordinates of the mice in the arena (Figure 1, left) under the effect of drugs from the three classes (Table 1), are first quantified using attributes—relevant movement variables (Figure 1, right). (2) a large

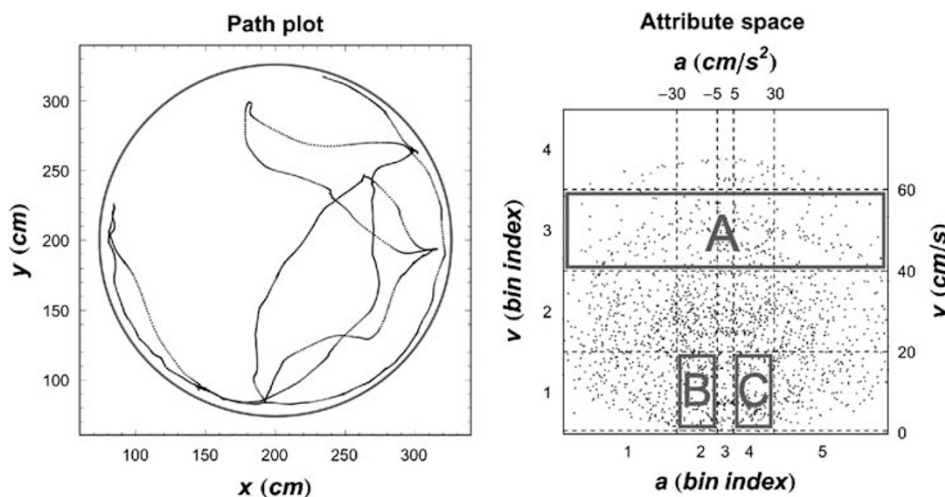


Figure 1 Path plot coordinates of a vehicle-injected mouse in the arena (left) and their representation in the attribute space (v, a) (right) (see 'Behavioral attributes' for details). Each point (v, a) in the attribute space corresponds to a data point (x, y) in the path plot. Dashed lines in the attribute space represent bin boundaries (see Table 2) dividing the space into 'cells' (patterns). Box A represents the pattern named $P\{**3,***,***\}$. Boxes B and C represent patterns $P\{**1,2,***,***\}$ and $P\{**1,4,***,***\}$, respectively. The relative frequency of using each pattern is measured as the number of data points in the corresponding 'cell' divided by the total number of progression data points in the session (see Figure 2).

number of behavior patterns are then defined as different combinations of attribute values (Table 2, Figure 1, right) and the frequency of using each pattern by each animal is measured (Figure 2); (3) a classification model is 'trained' by mining for patterns that best predict each psychopharmacological class (Figures 2 and 3); (4) to prevent

overfitting (ie patterns that were actually selected to fit idiosyncratic peculiarities of our particular data rather than general class properties) the above is conducted in the 'training data set' (Figures 5 and 7, left), and then only the discovered class predictors are tested in an independent 'test data set' measured in different animals (Figures 5 and 7, right); (5) finally, we validate our classification model by its ability to correctly classify additional drugs that it did not encounter during the training process (Figure 9)—a simulation of novel compound classification in drug discovery. In addition we estimate the heritability and replicability of the discovered behaviors in naive mice by measuring them in previously collected data from 10 inbred strains across three laboratories (Kafkafi *et al*, 2005; Kafkafi and Elmer, 2005).

Table 1 List of Drugs

Drugs	Doses (mg/kg)	Data set		
		Training	Test	Validation
<i>Psychomotor stimulants</i>				
Cocaine	3.0, 5.6, 10.0, 17.0, 30.0	X	X	
Methamphetamine	0.3, 1.0, 1.7, 3.0	X	X	
Methylphenidate	1.7, 5.6, 10.0, 17.0	X	X	
Mazindol	1.0, 3.0, 5.6, 10.0			X
Modafinil	30, 56, 100, 170			X
<i>Opioids</i>				
Morphine	1.0, 3.0, 5.6, 10.0	X	X	
Oxycodone	1.0, 3.0, 5.6	X	X	
Fentanyl	0.056, 0.1, 0.17, 0.3	X	X	
Codeine	5.6, 17.0, 30.0, 56.0			X
<i>Psychotomimetics</i>				
PCP	3.0, 5.6	X	X	
SDZ220851	1.7, 3.0, 5.6	X	X	
Ketamine	5.6, 10.0, 17.0	X	X	
Salvinorin A	1.0, 3.0, 10.0			X
<i>Mixture</i>				
Cocaine+morphine	10.0+5.6, 17.0+10.0			X

The drugs and doses used in the experiment and their division to classes. The training set was used to discover the best class predictors in nine drugs (third column). The test set was used for testing the predictors in these same nine drugs (fourth column). The validation set was used to test the predictors in drugs or combinations that were not used in predictor development (fifth column).

MATERIALS AND METHODS

Drugs, Animals, and Experimental Procedures

Drugs. In total, thirteen drugs representing three drug classes (opioid, psychotomimetic, and psychomotor stimulant) were investigated in this study (Table 1). Methamphetamine, cocaine, methylphenidate, mazindol, morphine, oxycodone, fentanyl, codeine, phencyclidine hydrochloride (PCP), salvinorin A, and modafinil were purchased from Sigma-Aldrich (St Louis, MO). SDZ220851 (SDZ) was purchased from Tocris (Ellisville, MO). Ketamine was purchased from Henry Schein (Melville, NY). All drugs except SDZ, mazindol, modafinil, and salvinorin A were dissolved in 0.9 % saline. SDZ was dissolved in Tween and ethanol, then brought to the appropriate concentration with vehicle (Tween (1%)/ethanol (5%)). Mazindol was dissolved in Tween, ethanol, and acetic acid, then brought to the appropriate concentration with vehicle (Tween (4%)/ethanol (3%)/acetic acid (1%)). Salvinorin A was dissolved in Tween, then brought to the appropriate concentration with vehicle (Tween (5%)). Modafinil was dissolved in Tween and ethanol, then brought to the appropriate concentration with deionized water vehicle (Tween (5%)/ethanol (5%)). Vehicle solutions were used as control. Morphine, oxycodone, fentanyl, codeine, and salvinorin A were given subcutaneously (s.c.); all other drugs were given

Table 2 The 10 Attributes of Movement Used in PA

Symbol	Attribute definition	Units	Bin indexes	Bin boundaries
<i>s</i>	Time from beginning of session	Min	1, 2, 3	0, 20, 40, 60
<i>d</i>	Distance from arena wall	cm	1, 2, 3, 4	0, 5, 15, 30, 125
<i>v</i>	Speed of movement	cm/s	1, 2, 3, 4	0, 20, 40, 60, ∞
<i>a</i>	Acceleration of movement	cm/s ²	1, 2, 3, 4, 5	−∞, −30, −5, 5, 30, ∞
<i>j</i>	Jerk (change in acceleration) of movement	cm/s ³	1, 2, 3, 4, 5	−∞, −300, −50, 50, 300, ∞
<i>h</i>	Movement direction relative to wall (heading)	Degree	1, 2, 3, 4, 5	−90, −30, −5, 5, 30, 90
<i>c₄</i>	Path curvature in a 4-cm scale (turning)	Degree/cm	1, 2, 3, 4, 5	−∞, −5, −1, 1, 5, ∞
<i>c₁₆</i>	Path curvature in a 16-cm scale (turning)	Degree/cm	1, 2, 3, 4, 5	−∞, −2.5, −0.5, 0.5, 2.5, ∞
<i>t_s</i>	Time from start of progression segment	s	1, 2, 3	0, 0.2, 1, ∞
<i>t_e</i>	Time to end of progression segment	s	1, 2, 3	0, 0.2, 1, ∞

These attributes are computed for each coordinate from the progression segment of the animal. The range of each attribute is divided into several bins, with indices and boundaries specified in the two rightmost columns.

intraperitoneally (i.p.). All drugs were given at an injection volume of 10.0 ml/kg except for salvinorin A (10 mg/kg dose) and modafinil (100 mg/kg dose); the latter drug doses were administered at an injection volume of 16.0 ml/kg due to solubility concerns.

Animals. Animals used in these experiments were all 60- to 80-day-old C57BL/6J male mice (Jackson Laboratories) allowed to acclimate to the animal facility at least 7 days before testing. They were housed five per cage in standard conditions of 12:12 light cycle, 22°C room temperature, water and food *ad libitum*. The experimental protocols followed the 'Principles of Laboratory Animal Care' (NIH publication no. 86-23, 1996). The animals used in this study were maintained in facilities fully accredited by the American Association for the Accreditation of Laboratory Animal Care.

For statistical reasons (see 'Multiple comparisons correction') animals in each dose group of nine drugs were approximately equally divided into a 'training set' and a 'test set' (Table 1) mostly including $n = 4$ animals each with a range of 3-7 except for the saline groups ($n = 8$ and 12 in the training and test set respectively; see 'Pattern mining' for the reason to enlarge the saline group). The SDZ vehicle ($n = 5$ and 4, respectively) did not produce statistically different significant results from the saline in any part of the analysis. Four additional drugs (Table 1) were used to validate the classification with n of 3-7 per dose group including their vehicle groups. The person running the animals and performing the analysis (NK) was blind to the identity of mazindol and codeine, whereas modafinil and salvinorin A were suggested by an anonymous reviewer. In the training and test sets, drugs and doses were assigned so that in the same cage drugs from all three classes or two classes and vehicle were used, and no two animals received the same drug. Overall a total of 347 animals were used in this study.

Testing procedures. Open-field tests were conducted during the light phase of the cycle using the standard procedure in SEE (Software for the Exploration of Exploration; see Draï and Golani, 2001; Kafkafi *et al*, 2003a; Kafkafi *et al*, 2005; Kafkafi and Elmer, 2005). Briefly, each animal was injected once and immediately introduced to a 2.50 m diameter circular open-field arena for 60 min while its location was tracked using Noldus EthoVision video-tracking system at a rate of 30 Hz and the highest video resolution (480 × 640 pixel). The {Time, X, Y} coordinates of the path (eg Figure 1, left) were exported to SEE, and the SEE Path Smoother procedure (Hen *et al*, 2004) was used to filter out tracking noise. Each session file thus included 30 coordinates × 60 s × 60 min = 108 000 data points. Using the standard SEE procedure the path was further divided into segments of progression and lingering (small local movements during stopping; see Draï and Golani, 2001; Kafkafi *et al*, 2001). Current PA analysis utilizes only the progression component of the behavior, as the spatial resolution of the lingering component is too low to be reliably measured by current tracking technology. Progression segments included on average 76 400 data points per session with an SD of 15 500. Four inactive animals with less than 29 250 data points each (three

standard deviations below the mean) were discarded from the analysis.

Previously collected data. To estimate the broad-sense heritability and replicability of the discovered behaviors we measured them in the open-field behavior of naive mice from 10 inbred strains across three laboratories, recorded in a previous study. The test protocol of that study, detailed in Kafkafi *et al* (2005) and Kafkafi and Elmer (2005), differed from the protocol of the present study in two principal ways: the mice were not injected, and the session duration was 30 min rather than 60 min. Inbred strains included the common A/J, BALB/cByJ, C3H/HeJ, C57BL/6J, DBA/2J, FVB/NJ, SJL/J, and 129S1/SvImJ, as well as the wild-derived CAST/EiJ and CZECHII/EiJ. The three laboratories replicating the experiment were the National Institute of Drug Abuse—IRP in Baltimore, the Maryland Psychiatric Research Center (MPRC) in the University of Maryland, and the Department of Zoology in Tel Aviv University. The present study was also conducted in the MPRC, but using new animal facilities and arena in a new building, so for most practical purposes it can be considered a different laboratory.

Pattern Array Analysis

PA is a method developed for analyzing rodent path data and mining it for patterns that best predict any experimental factor of interest (ie genotype, drug etc). The PA algorithm was described in detail in Kafkafi *et al* (in press) where it was used to discover early symptoms in an animal model of Amyotrophic Lateral Sclerosis (SOD1 transgenic rats) at approximately half the age it is diagnosed by the standard behavioral tests. Here we describe it again while focusing on the application of PA to psychopharmacological drug prediction and the few adjustments needed for its adaptation from rats to mice. As in other data mining strategies, PA tests many different hypotheses in parallel, in this case testing the difference between the classes in each of $\sim 10^5$ different behavioral patterns. Hence, the main problem it faces is how to dissect the behavior into multiple relevant categories.

Behavioral attributes. The first step of PA analysis is representing each path coordinate out of the progression of the animal using several attributes (using data mining terminology, see Tan *et al*, 2006), which in our case are dynamical variables of movement such as the momentary distance from the wall, the momentary speed of movement, the momentary acceleration, the momentary direction of movement, and the momentary change in direction (turning). The 10 attributes chosen for this study (Table 2) consist of the variables found most useful for the analysis of open-field behavior throughout many previous studies. The distance from the wall d was shown to measure heritable thigmotactic behavior (Broadhurst, 1975; Leppanen *et al*, 2005; Lipkind *et al*, 2004) highly replicable across laboratories (Kafkafi *et al*, 2005). The momentary speed v was shown to be a key variable in the intrinsic categorization of behavior into progression and 'lingering' in both mice (Draï and Golani, 2001) and rats (Kafkafi *et al*, 2001). The acceleration a was shown to have high heritability and

replicability in inbred mouse strains (Kafkafi *et al*, 2003b). The jerk j (the derivative of acceleration according to time, or the second derivative of the speed) was chosen because speed peaks were shown to be a meaningful component of rodent behavior (Drai and Golani, 2001; Kafkafi *et al*, 2001) and j is required to distinguish between speed peaks (near-zero acceleration and negative jerk) and local minima of speed (near-zero acceleration and positive jerk). The momentary heading h (direction of movement relative to the arena wall) was proposed by Horev *et al* (2007) as an important aspect of open-field behavior in the mouse. The path curvature in a scale of 4 and 16 cm (c_4 and c_{16} , respectively) were shown to discriminate several mouse inbred strains with high heritability and replicability (Kafkafi and Elmer, 2005). In this study h , c_4 , and c_{16} are all defined as positive in a direction away from the arena wall and negative toward the wall. The time from the beginning and to the end (t_s and t_e , respectively) of progression segments (Drai and Golani, 2001; Kafkafi *et al*, 2001) make it possible to mine patterns that always occur, eg when initiating a movement or when stopping. To these nine attributes that were also used in rats (Kafkafi *et al*, in press) we add a tenth attribute in this study: the time s from the beginning of the session. This attribute was added because some of the drugs in this study may differ from others in their rate of onset.

Every path coordinate from the progression of the animal can thus be represented using a list of 10 attributes of the form ($s, d, v, a, j, h, c_4, c_{16}, t_s, t_e$). However, in practice we only consider up to any four attributes at any given time (see below), and mark the irrelevant attributes with asterisks (standing for 'wildcards', ie attributes that can accept any value). For example, using only the speed v and acceleration a is represented by a partial list of the form ($*, *, v, a, *, *, *, *, *$). Any partial list of attributes can be graphically represented as constructing an 'attribute space', in which each path coordinate is represented by a point. For example, Figure 1 (right) shows the path of the animal in the space (in this case a two-dimensional plane) defined by the two attributes a and v .

Attribute partition. The range of each attribute is partitioned into several bins indexed 1, 2, 3, ... (Table 2), eg the attribute of distance from the wall d is partitioned into bin 1: 0–5 cm (approximately maintaining physical contact with the wall), bin 2: 5–15 cm (close but not touching), bin 3: 15–30 cm (moderately away from the wall), and bin 4: > 30 cm (far into the center). The number of bins and their boundary values are chosen to reasonably fit the typical range and achieve an approximately equal distribution; eg we use bins of increasing width for d because rodents spend a much larger portion of the session near the wall. Figure 1 (right) depicts equal partition into bins (dashed lines) in the speed attribute v (vertical axis) and unequal division in the acceleration attribute a (horizontal axis). The number of bins in this study is the same as was used in Kafkafi *et al* (in press) for rats, but the boundary values in some of the attributes were slightly modified to fit the typical attribute range and distribution in mice.

Definition of patterns. PA can now define multiple patterns of movement as combinations of bins from one or more attributes. We code these patterns using the same order as

in the list of attributes ($s, d, v, a, j, h, c_4, c_{16}, t_s, t_e$) by denoting the bin index in each attribute (see Table 2) and using asterisks to denote attributes that can accept any value and are therefore irrelevant to the definition of this pattern. For example, the single-attribute pattern coded $P\{*, *, 3, *, *, *, *, *, *, *\}$ is defined only by the third bin ($40 < v < 60$ cm/s) of the third attribute (speed), ie the animal is moving moderately fast. This pattern can be graphically described as the 'cell' marked A in the attribute space in Figure 1 (right). The two-attribute pattern-coded $P\{*, *, 1, 2, *, *, *, *, *, *\}$ is defined as the first bin in the third attribute and the second bin in the fourth attribute, ie that the animal is moving very slowly ($0 < v < 20$ cm/s) and is also slightly decelerating ($-30 < a < -5$ cm/s²). This pattern can be graphically represented by cell B in Figure 1. As more attributes are added to the definition of a pattern it becomes more and more specific, eg the four-attribute pattern $P\{*, *, 1, 2, *, 1, 5, *, *, *\}$ means moving very slowly while slightly decelerating in the direction of the arena wall but turning sharply away from it. However, as in Kafkafi *et al* (in press) we do not use patterns of more than four attributes because this would amount to an astronomical number of combinations, most of them so overspecified that they rarely occur in normal behavior. Using all possible bin combinations in up to any 4 out of the 10 attributes amounts to a total of 73 042 different behavior patterns tested in this study, which is more than the 50 674 patterns tested in Kafkafi *et al* (in press) due to the addition of attribute s .

Pattern mining. The difference between the experimental groups is then tested using the relative frequency of performing each of the 73 042 patterns as the dependent variable. The relative frequency of a pattern is computed as the time the animal spent in this pattern divided by the total progression time of the animal during the session. For example, the relative frequency of using each of the patterns in Figure 1 (right) is the number of points in the corresponding cell divided by the total number of points, and is expressed in percent as the heights of columns in Figure 2. For the needs of the statistical analysis this frequency is expressed using the logit transformation, commonly used to transform ratios. The test used in this study is simply a t -test of all the animals in each class of drugs *vs* all the animals in the other classes and vehicle groups, pooled over drug and dose. This is illustrated in Figure 2 for three animals injected with psychomotor (top row) *vs* three animals injected with other drugs (bottom row). The vehicle group was larger than the other dose groups in order to increase the power to discover class predictors that identify difference from vehicle as well as from other classes. PA is equally amenable to more elaborate tests appropriate for detecting the effect of drug and dose, but this is outside the scope of the present paper.

Multiple comparisons correction. As with similar parallel data mining strategies, the quandary with such an approach is the multiplicity problem. That is, when using a significance level of 0.05 we expect 1 out of 20 patterns to be 'significant' just by pure chance. When simultaneously screening a large number of possible patterns we thus need to prohibit the occurrence of false positives and provide valid statistical inference for the selected patterns

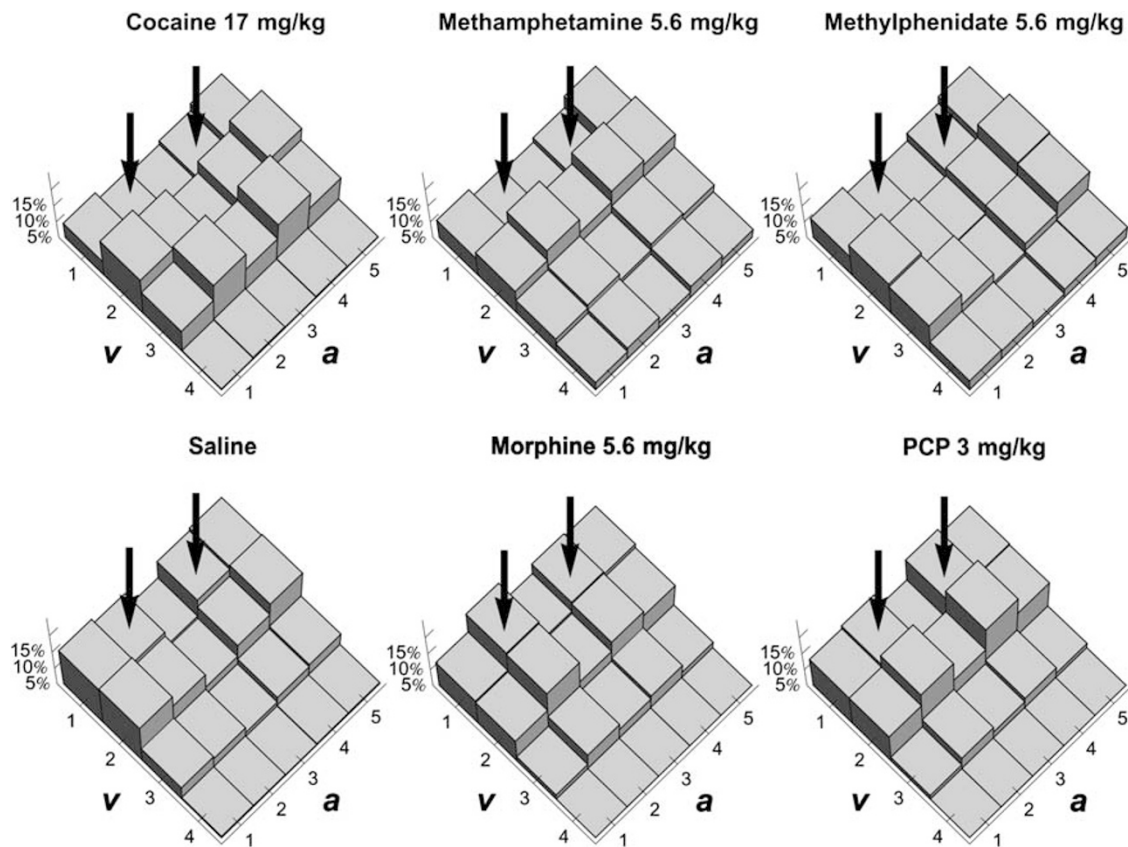


Figure 2 Pattern arrays of six animals out of the data. All graphs are three-dimensional histograms showing the frequency (horizontal axis) of the 20 patterns in the $(**v,a,*****)$ attribute combination of speed v and acceleration a , with bin indices (1–4 for a , 1–5 for v) as shown in Table 2. The patterns $P\{*,*,1,2,*,*,*,*,*\}$ and $P\{*,*,1,4,*,*,*,*,*\}$ (B and C in Figure 1) are marked by the left and right arrows respectively in each animal. The relative frequency of these two patterns was found by Pattern Array (PA) to significantly decrease under the effect of psychomotor stimulants, (top row) but not under the effect of other drugs (bottom row). See Figure 8 for actual instances of these two patterns.

(Benjamini and Yekutieli, 2005). As in Kafkafi *et al* (in press) we protect ourselves in two ways: first by applying the strictest multiple comparisons criterion, the Bonferroni criterion, which uses a corrected significance level of α/n , where n is the number of comparisons, thus ensuring that the probability of discovering even a single false pattern is less than α . The comparison of 73 042 different patterns using $\alpha = 0.05$ yields a Bonferroni criterion as $p < 0.05/73\ 042 = 6.8 \times 10^{-7}$. Second, as is recommended for data mining studies (Tan *et al*, 2006) we divide our data into two independent sets—the training set and the test set (see ‘Drugs, Animals, and Experimental Procedures’ and Table 2). The mining process described above is performed in the training set, and only those patterns found significant using the Bonferroni are then tested in the test set. In addition to these two steps we evaluate the discovered patterns by their ability to classify several additional drugs (Table 1, last column) that were not used at all in the process of isolating these patterns.

Heritability and Replicability Analysis

We used previously collected data in naive mice from 10 inbred strains across three laboratories (see ‘Previously collected data’) to statistically estimate the broad-sense heritability and the replicability of the discovered class

predictors. This analysis was conducted using mixed-model ANOVA of Genotype (fixed variable) \times Laboratory (random variable), calculated in S-PLUS 2000 statistical software (Insightful, Seattle) with the linear mixed effects function for significance estimation and the REML (restricted maximum likelihood) method for percentage of variance estimation. This is the same procedure utilized in Kafkafi *et al* (2005) and Kafkafi and Elmer (2005) with previous SEE endpoints in these same animals. In this procedure the broad-sense heritability is calculated as the percentage of total variance attributed to the genotype by the mixed ANOVA. This is in fact a conservative estimation of heritability, because it completely excludes variance due to interaction with the laboratory (interaction that is likely to include some genetic component), and because it employs the conservative mixed model and REML methods. For extensive discussion of the mixed model approach to replicability as well as detailed explanation of these procedures see Kafkafi *et al* (2005).

RESULTS

Mining for Opioid Predictors

First we use PA with *t*-test to test all opioid treated animals (morphine, oxycodone, and fentanyl) *vs* all the other

animals in the training set, pooled over drug and dose. In 19 260 out of the total 73 042 patterns this difference was found more significant than $p = 6.8 \times 10^{-7}$ (the Bonferroni criterion at a level of $\alpha = 0.05$). Such a high percentage (more than 26%) of significant patterns is an indication that the overall effect of the opioids was very strong. Figure 3 (left) shows only those patterns more frequent than 1% in behavior. In fact, as Figure 3 (left) shows, many of those patterns were more significant than the Bonferroni criterion by several orders of magnitude (points high above the dashed line). Out of the most significant patterns we prefer to use those that are also more frequent in behavior, as they are likely to be more reliable. Therefore Figure 3 plots the significance of patterns *vs* their mean relative frequency in all animals of the training set, regardless of class. We chose pattern $P\{*,*,*,4,*,*,*,*\}$ (marked by an arrow in Figure 3, left) as our 'opioid predictor' although it was not the most significant pattern, because it was very abundant in behavior, more than 20% of the total progression time on average. This is a single-attribute pattern in which the fifth attribute, the jerk j (change in acceleration), accepts an index of 4, defined as moderately positive jerk ($50 < j < 300 \text{ cm/s}^3$; see Table 2). This means that the animal is increasing its acceleration, possibly switching from braking to 'speeding up'.

Figure 4 illustrates some particular occurrences of this pattern out of the data of four animals in the training set, representing control behavior and each of the three classes. The acceleration a is indicated by the slope of the speed profile—negative ('braking down') when it descends and positive (speeding up) when it ascends. The jerk j is indicated by any change in this slope, with the pattern $P\{*,*,*,4,*,*,*,*\}$ defined as a moderate positive change. Such a change mainly includes moderate switches from descending to ascending slope (ie like switching from the brakes to the gas pedal while driving) that are shown as local minima in the speed profile. However, this pattern also includes descending slopes moderately becoming less

descending ('gently releasing the brake pedal') or ascending slopes moderately becoming more ascending ('gently stepping on the gas'). Note that at the 5.6 mg/kg oxycodone dose the speed of the animal is more regular than under saline or nonopioid drugs, with progression segments typically appearing almost rectangular in the speed profile, which decreases the frequency of positive jerk.

Figure 5 (left) shows that all the opioid drugs consistently decreased the frequency of the opioid predictor $P\{*,*,*,4,*,*,*,*\}$ in the training set, seen on the vertical axis (the horizontal axis in this figure represents the psychomotor stimulant predictor, which is discussed in 'Mining for Psychomotor Predictors'). The size of the decrease was not large, from about 25% in control, psychomotor, and psychotomimetic animals down to about 20% or slightly less in high opioid doses. However, this decrease was very significant, representing about six or more standard deviations of the saline animals. Note that in Figure 5 (as well as in Figures 7 and 9) the pattern frequencies in percents of the total progression time uses the same units as the frequency in Figure 3 horizontal axis, but in Figure 3 it is the mean frequency over all animals in the training set regardless of class, drug and dose, whereas Figures 5, 7 and 9 show pattern frequency of particular drugs and doses. Note also that the decrease in the frequency of the opioid predictor in the opioids drugs is generally dose dependent, which is of particular interest as the patterns in this study were not explicitly selected for dose dependency, but identified based on distinguishing the class effect regardless of dose. The psychomotor and psychotomimetic drugs did not have a significant effect on the opioid predictor.

The effect of the chosen opioid predictor was validated in the test set (Figure 5, right) and again it was significantly decreased by the opioid drugs ($F_{1,103} = 75.3$, $p = 6.7 \times 10^{-14}$). As the test set was used only for testing the few patterns chosen in the training set, it does not require correction for the comparison of 73 042 different patterns.

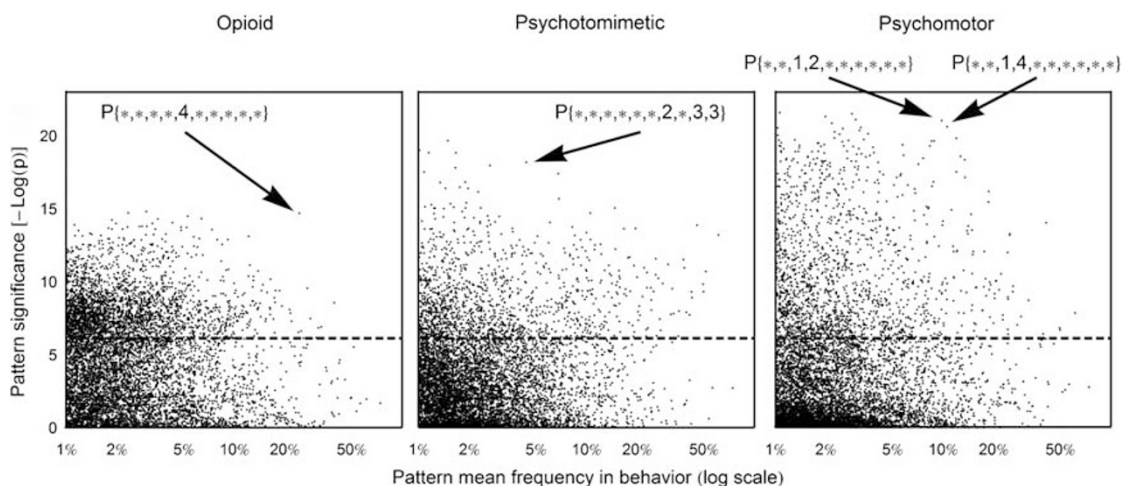


Figure 3 The significance of multiple movement patterns (represented as dots) when mined for opioid prediction (left), psychotomimetic prediction (middle), and psychomotor stimulant prediction (right). Pattern significance (vertical axes) is presented as a function of the pattern frequency in behavior (horizontal axes). Pattern frequency is calculated as its mean percentage out of the total progression time over all animals in the training data set, using a log scale. Only patterns more frequent than 1% are shown. Pattern significance is measured by the p -value order of magnitude $-\log_{10}(p)$, with dashed lines indicating the Bonferroni criterion $\log_{10}(-6.8 \times 10^{-7}) = 6.17$, hence patterns above them are statistically significant. Arrows identify the patterns chosen as class predictors (see text and Figures 4, 6, and 8) by their Pattern Array (PA) code names.

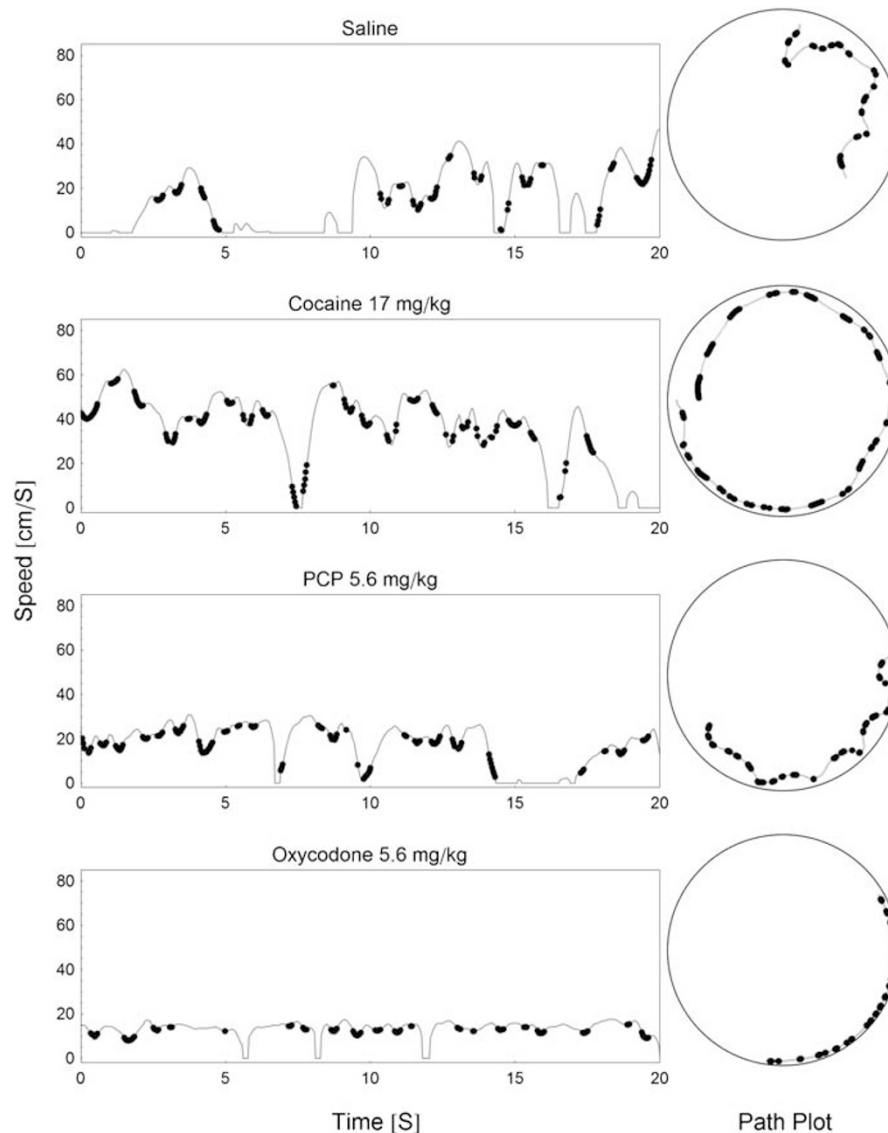


Figure 4 Particular occurrences of the opioid predictor pattern $P\{*,*,*,4,*,*,*,*\}$ ('moderately positive jerk') as performed by four animals (rows) out of the training data set, injected by saline and three drugs representing the three classes. For each animal 20 s of behavior are shown both in the speed profile (left) and in the path plot (right), with data points belonging to the opioid predictor pattern are bolder. Note that the speed profiles show progression segments bracketed by stops (0 speed) that were not included in the analysis. The mouse injected by the opioid drug oxycodone uses this pattern less out of its progression time. See text for further explanations.

However, note that this p -value is much more significant even than the Bonferroni criterion, ie even if we were to switch the training and test data sets, our chosen opioid predictor would still be found significant. The test set results also show similar dose dependency as the training set results, and the frequency values themselves are similar as well for most drugs and doses.

Mining for Psychotomimetic Predictors

Next we test all psychotomimetic-treated animals (PCP, SDZ, and ketamine) vs all other animals in the training set, pooled over drug and dose, in the same procedure used for isolating the opioid predictor patterns. In total, 10 518 patterns passed the Bonferroni criterion, many of them by several orders of magnitude (Figure 3, middle). As with the opioid predictor, we search for a psychotomimetic

predictor pattern that is not only very frequent in behavior or only very significant, but a combination of the two. Unlike the case with the opioids (Figure 3, left) here the best compromise was not so obvious. While we chose $P\{*,*,*,*,*,2,*,3,3\}$ as our psychotomimetic predictor, choosing any other of the reasonable compromises between frequency and significance in Figure 3 (middle) gives essentially similar results.

As can be inferred from Table 2, $P\{*,*,*,*,*,2,*,3,3\}$ is a three-attribute pattern defined as slight turn in the direction of the arena wall ($-5 < c_4 < -1^\circ$ per cm, with the minus sign representing wall direction) during the middle of progression segments (at least 1 s after the beginning of movement and at least 1 s before stopping). In Figure 6 specific occurrences of this pattern are highlighted in the same sections of behavior shown in Figure 4. Note in the path plots that under the psychotomimetic drug PCP the path

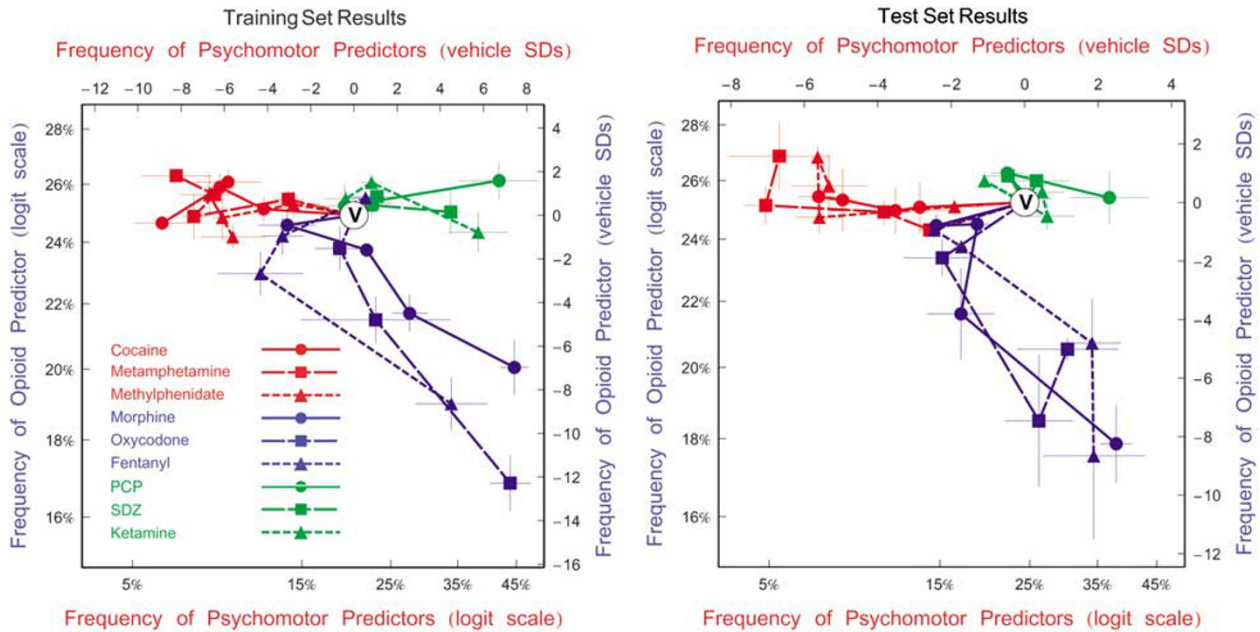


Figure 5 Separation of drugs into classes in the training data set (left) and in the independent test data set (right). The results are shown in the plane of the psychomotor stimulant predictor patterns frequency on the horizontal axis vs the opioid predictor pattern frequency on the vertical axis. Frequencies are given in two different units: either in percent out of the total progression time in the session (bottom and left axes) or in SDs of the vehicle group (top and right axes). Each series shows a drug's dose-response, with the doses given in Table 1 connected in ascending order, and symbols representing dose group means ± SEs. Series in red: psychomotor stimulant drugs. Series in blue, opioid drugs; series in green, psychotomimetic drugs; V, vehicle mean.

of the animal includes more turning. Interestingly, however, the same significant difference was not found in turning away from the wall. Looking at the results of $P\{*,*,*,*,*,4*,3,3\}$ (ie the same as $P\{*,*,*,*,*,2*,3,3\}$ except with turning away from the wall instead of toward it) reveals that animals under the effect of psychotomimetics turn more than the control animals both toward the wall and away from it. However, opioids also cause some increase in turning away from the wall, which is why $P\{*,*,*,*,*,4*,3,3\}$ is not a significant psychotomimetic predictor, whereas $P\{*,*,*,*,*,2*,3,3\}$ is. That is, the way PA was applied in this study it isolates the patterns that best differentiate a class from the other classes in the study, rather than the patterns that characterize a class in itself. This example illustrates the use of PA for exploratory data analysis and further investigation of the behavior.

In addition, as Figure 6 suggests control animals turn in all directions nearly as much as psychotomimetic animals, but they frequently do it in the beginning or end of progression segments, whereas psychotomimetic animals mainly turn in the middle of progression segments. It thus appears that the natural turning of the control animals is usually a response to the same stimulus that made them start moving or stop, whereas the turning caused by psychotomimetic drugs is usually an intrinsic effect of the drug rather than a response to external stimuli.

Figure 7 (left, horizontal axis) shows that the chosen psychotomimetic predictor $P\{*,*,*,*,*,2*,3,3\}$ increased about twofold in all the psychotomimetic drugs in the training set, from 5% in the vehicle to about 10% in the high doses. This increase represents about 3–6 standard deviations of the vehicle animals. Again the increase is dose

dependent although the predictor was not explicitly selected for dose effect. Psychomotor stimulant and opioid drugs leave the psychotomimetic predictor unchanged or decrease it up to twofold and more. The effect of the psychotomimetic predictor was validated in the test set (Figure 7, right) and again it was significantly increased by psychotomimetics drugs ($F_{1,103} = 60.1, p = 6.5 \times 10^{-12}$). As with the opioid predictor pattern, this p -value is much more significant even than the Bonferroni criterion, implying that this pattern would still be found significant if we were to switch the training and test data sets. The test set results also show generally reliable replication of dose dependency and absolute values for most drugs and doses. Using the plane of the psychotomimetic predictor vs the opioid predictor in Figure 7 produces clear separation of opioids from psychotomimetics. Interestingly these predictors also separate the psychomotors from both the opioids and psychotomimetics, although they were not selected for it in any way.

Mining for Psychomotor Stimulant Predictors

We now test all psychomotor stimulant-treated animals (cocaine, methamphetamine, and methylphenidate) vs all the other animals in the training set, pooled over drug and dose, in the same procedure used for isolating the opioid and psychotomimetic predictor patterns. Altogether, 10070 patterns passed the Bonferroni criterion at a level of $\alpha = 0.05$, many of them by several orders of magnitude (Figure 3, right). Here we identify two patterns, $P\{*,*,1,2*,*,*,*,*,*\}$ and $P\{*,*,1,4*,*,*,*,*,*\}$ (marked by arrows in Figure 3, left) that constitute a good compromise

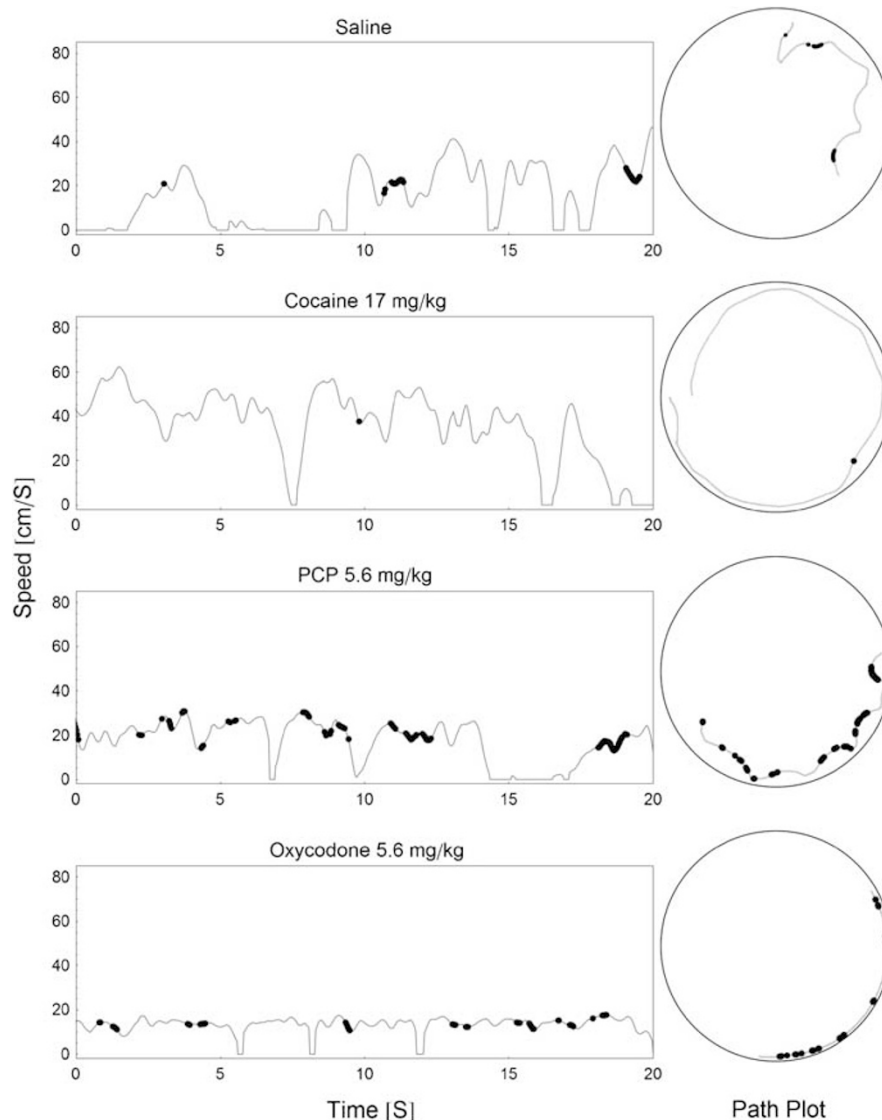


Figure 6 Particular occurrences of the psychotomimetic predictor pattern $P\{*,*,*,*,2,*,3,3\}$ ('turning toward the wall in the middle of progression segments') as performed by four animals (rows) out of the training data set, injected by saline and three drugs representing the three classes. For each animal 20 s of behavior are shown both in the speed profile (left) and in the path plot (right). These behavior sections are the same as those shown in Figure 4, but with data points belonging to the psychotomimetic predictor pattern are bolder. The mouse injected by the psychotomimetic drug phencyclidine hydrochloride (PCP) uses this pattern more often. See text for further explanations.

between significance and frequency. These two patterns are also illustrated in Figure 1, right (cells A and B) and in Figure 2 (columns mark by arrows). As these two patterns are explicitly nonoverlapping, as shown by the different indices in the forth attribute, we chose their union as our psychomotor predictor.

According to Table 2, $P\{*,*,1,2,*,*,*,*,*\}$ is a two-attribute pattern defined as moving slowly ($v < 20$ cm/s) while slightly braking ($-30 < a < -5$ cm/s²). The symmetric pattern $P\{*,*,1,4,*,*,*,*,*\}$ is defined as moving slowly while slightly speeding up ($5 < a < 30$ cm/s²). The union of these two patterns thus means moderately changing slow progression. As Figure 8 suggests, animals under the effect of the psychomotor drugs tend to move fast, using the low speed range only briefly when strongly accelerating in the beginning of movement or strongly decelerating when they

stop. Figure 5 (left, horizontal axis) shows that all psychomotor drugs decreased the frequency of the psychomotor predictors in the training set by more than twofold, from more than 20% in the control to less than 10% in the high doses, which represents approximately 6 standard deviations of the vehicle group. The decrease was again largely dose dependent. Generally using the plane of the psychomotor predictor patterns vs the opioid predictor pattern very clearly separates the high doses of psychomotor stimulants and opioids, and also separates the psychotomimetics from either psychomotors or opioids. However, the separation between low doses of psychomotors and opioids is weak, and in fact it appears that opioids behave very much like psychomotors in low doses before they change direction in the high doses. This similarity was also common to many other highly significant patterns, and

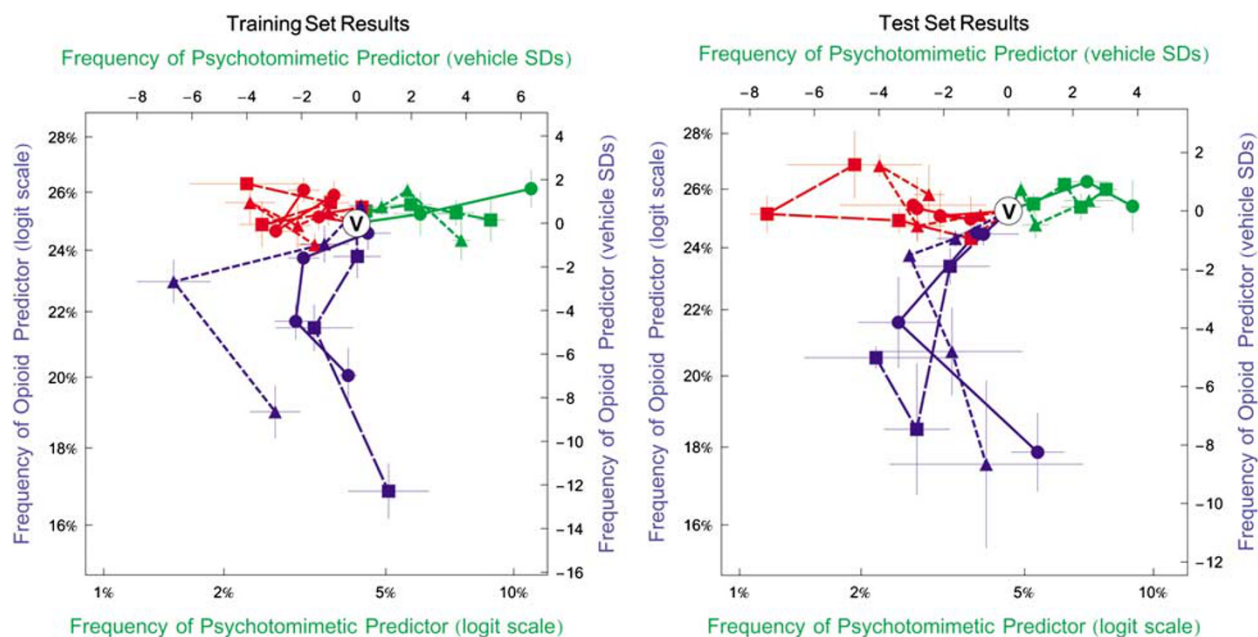


Figure 7 Separation of drugs into classes in the training data set (left) and in the independent test data set (right) is shown in the plane of psychotomimetics predictor pattern frequency on the horizontal axis vs the opioid predictor pattern frequency on the vertical axis. Units, colors, and symbols are as in Figure 5.

suggests strong similarity of open-field behavior under these low doses of psychomotors and opioids, or that the difference is in attributes other than those chosen for the present study. The psychomotor predictor patterns were validated in the test set (Figure 5, right) and again were very significant either when using their union ($F_{1,103} = 84.7$, $p = 3.2 \times 10^{-15}$) or in each in itself ($F_{1,103} = 86.4$, $p = 1.9 \times 10^{-15}$ for $P\{*,*,1,2,*,*,*,*,*,*\}$, $F_{1,103} = 90.6$, $p = 1.9 \times 10^{-16}$ for $P\{*,*,1,4,*,*,*,*,*,*\}$), mostly with dose dependency and particular dose values reliably replicated.

Classifying 'Unknown' Compounds

We evaluated the PA classification model by its ability to classify four additional drugs (Table 1, last column) that were not used in the isolation of these predictors. This is a simulation of classifying novel compounds with unknown psychopharmacological properties in the process of drug discovery. Mazindol and codeine were introduced by one of the authors (GE) with the author testing the animals and running the analysis (NK) masked to their identity, class, and dose. Modafinil and salvinorin A were suggested by an anonymous reviewer. As shown in Figure 9, all four drugs were classified in a dose-dependent manner. The psychomotor stimulant mazindol produced a significant ($F_{4,17} = 3.3$, $p < 0.05$, using one-way ANOVA) psychomotor-like response in the psychomotor predictor (Figure 9, left). It also had a psychomotor-like trend in the psychotomimetic predictor (Figure 9, right). The psychomotor stimulant modafinil produced a smaller effect in range but still significant ($F_{3,12} = 8.4$, $p < 0.01$, one-way ANOVA) psychomotor-like response (Figure 9, left) and much smaller yet still significant ($F_{3,12} = 4.2$, $p < 0.05$, one-way ANOVA) opioid response, slightly separating it from the other psychomotors in this study. The opioid codeine produced a strong ($F_{4,15} = 11.0$, $p < 0.001$, one-way ANOVA)

opioid-like response in the opioid κ predictor, and like the previously tested opioids also a minor ($F_{4,15} = 3.2$, $p = 0.055$, one-way ANOVA) psychomotor response in low doses (Figure 9, left). Salvinorin A, a κ -agonist opioid with psychotomimetic properties, posed a particular challenge because the classification model was not trained with any κ -agonists. It had a marginally significant ($F_{3,22} = 2.9$, $p = 0.06$, one-way ANOVA) response in the psychotomimetic predictor (Figure 9, right) and also a trend ($F_{3,22} = 2.2$, $p = 0.11$, one-way ANOVA) toward opioids in low doses, especially in comparison with the other psychotomimetics drugs in the study. However, salvinorin A also showed a significant ($F_{3,22} = 7.3$, $p < 0.01$, one-way ANOVA) psychotomimetic-like response in the psychomotor predictor (Figure 9, left), illustrating the advantage of classifying a novel compound with a combination of several class predictors.

In addition we also evaluated the PA classification with mixtures of cocaine and morphine in the same injection. These mixtures score intermediate results between the opioids and psychomotor stimulant (Figure 9). Interestingly the mixture of 10 mg/kg cocaine with 5.6 mg/kg morphine scored similar to 10 mg/kg cocaine alone in the psychomotor predictor and very similar to 5.6 mg/kg morphine alone in the opioid predictor (dashed line in Figure 9, left). This suggests that at these doses the psychomotor predictor patterns are independent of the opioid predictor pattern both at the neuropharmacological level and the behavioral level. When increasing the doses to 17 mg/kg cocaine with 10 mg/kg morphine, however, the mixture scored considerably less than either 17 mg/kg cocaine alone in the psychomotor predictor or 5.6 mg/kg morphine alone in the opioid predictor. This suggests that at these higher doses there was interference between opioid and psychomotor activities at the neuropharmacological or behavioral level or both.

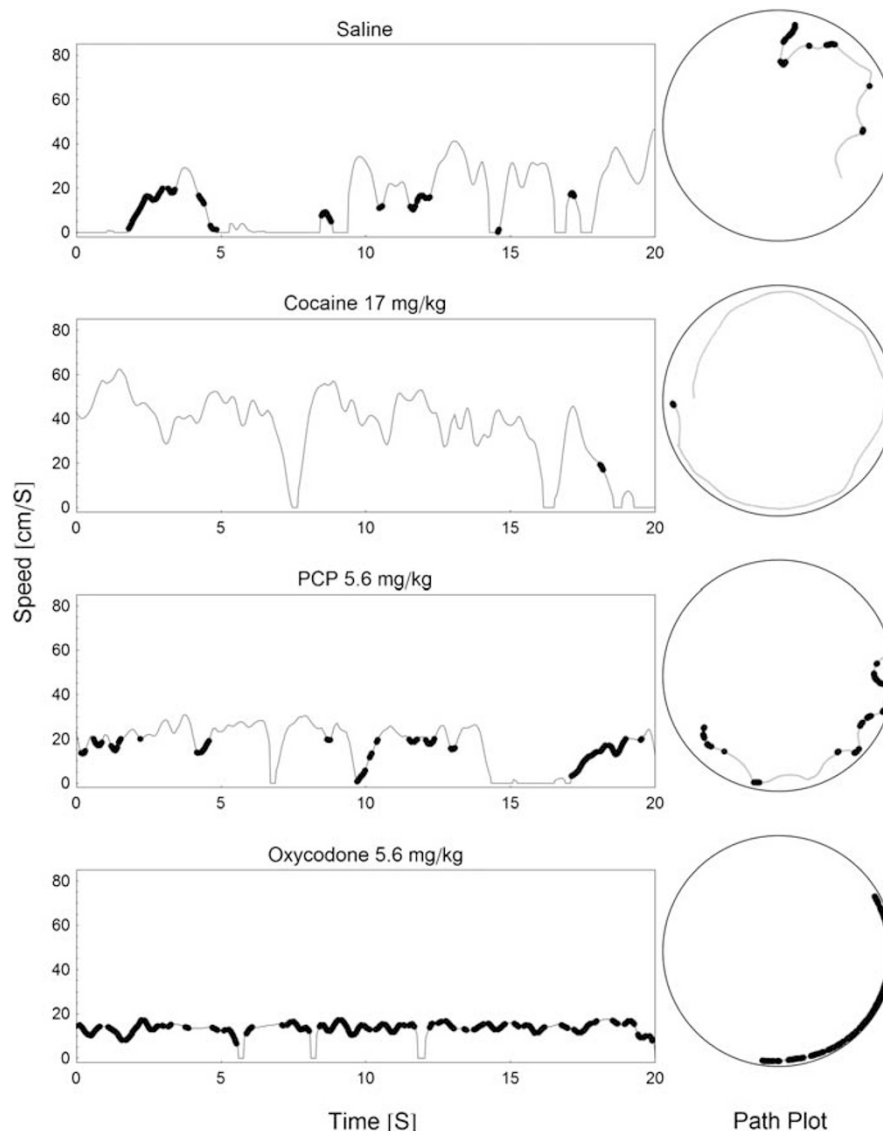


Figure 8 Particular occurrences of the psychomotor stimulant predictor pattern, the union of $P\{^{**},1,2,^{***},^{**},^{**}\}$ ('moderately braking while moving slowly') and $P\{^{**},1,4,^{***},^{**},^{**}\}$ ('moderately accelerating while moving slowly'), as performed by four animals (rows) out of the training data set, injected by saline and three drugs representing the three classes. For each animal 20 s of behavior is shown both in the speed profile (left) and in the path plot (right). These behavior sections are the same as those shown in Figure 4, but with data points belonging to the psychomotor predictor pattern are bolder. The mouse injected by the psychomotor stimulant cocaine uses this pattern much less. See text for further explanations.

Heritability and Replicability of Class Predictors

An important advantage of the database approach is that once new behavioral measures are isolated, they can readily be measured in stored data from previous experiments in diverse genotypes and conditions (Kafkafi and Elmer, 2005). We measured the discovered class predictors in naive mice data from 10 inbred strains across three laboratories to estimate their broad-sense heritability and the replicability, utilizing the mixed model approach (Kafkafi *et al*, 2005). Strain differences in all predictors were significant ($p < 0.0001$) despite the higher benchmark employed by the mixed model. The heritability of the opioid, psychotomimetic, and psychomotor predictors was 52.3, 59.6, and 58.9% respectively. Note that this is a conservative estimate (see 'Heritability and Replicability Analysis'). In comparison, the variance attributed to the interaction of genotype

with the laboratory was 14.4, 3.6, and 6.0% respectively. One possible issue with this comparison is that the naive mice data were measured in 30 min sessions, as opposed to 60 min sessions used in the present study. To address this issue we repeated the complete PA analysis of the drugs data using the first 30 min of each session only. The results were almost identical to the results of the 60 min analysis, except that the p -values were less significant by about two orders of magnitude, but still considerably more significant than the Bonferroni criterion.

DISCUSSION

The objective of this report was to explore a data mining approach for classifying a wide range of psychoactive drugs. Our working hypothesis was that psychoactive drugs

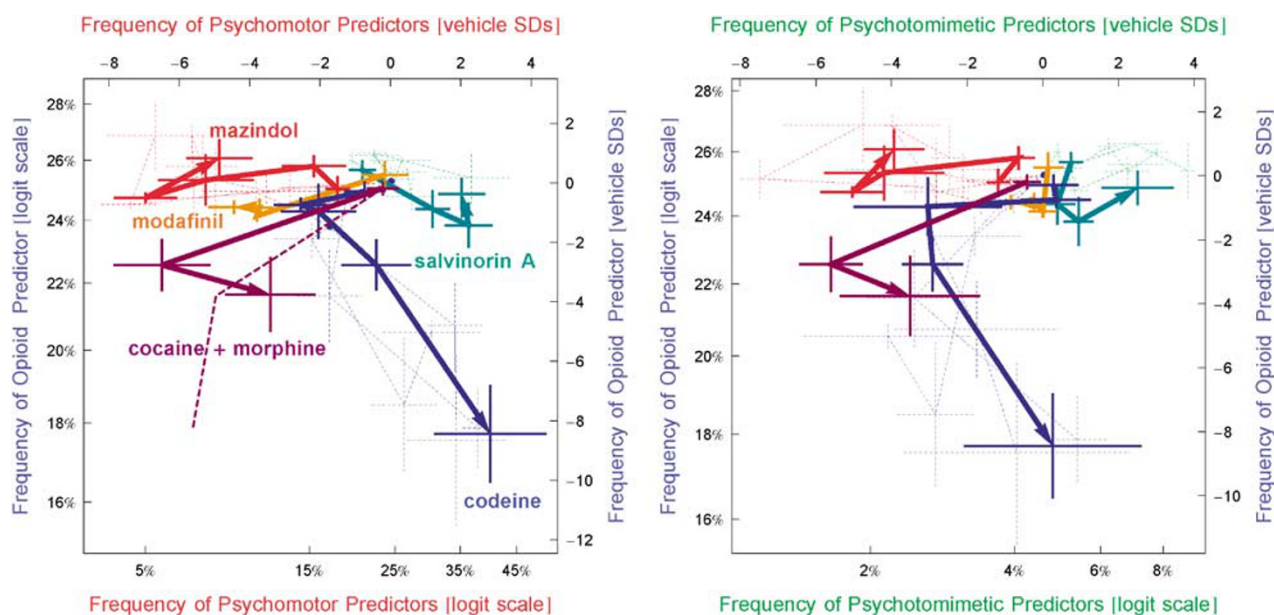


Figure 9 Using the discovered class predictors to classify mazindol (bold red series), modafinil (bold orange series), codeine (bold blue series), salvinorin A (bold cyan series), and mixture of cocaine and morphine (bold purple series). Each series depicts a dose–response, with arrows indicating increasing dose and dose values detailed in Table 1. Results are shown in the plane of psychomotor stimulant predictors frequency vs opioid predictor frequency (left), and in the plane of psychotomimetic predictor frequency vs opioid predictor frequency (right), superimposed over the test set drug results (Figures 5 and 7, right) in dotted red, blue and green. Purple dashed line (left) shows the effect of the cocaine–morphine mixture that would be expected from the same doses of cocaine alone and morphine alone. All other symbols and units are as in Figure 5.

possess psychopharmacological profiles that can be uniquely characterized when the behavioral resolution is sufficiently magnified. To this end we developed PA to mine multiple patterns of spatial behavior and discover those that best predict a class of psychoactive drugs (class predictors). A total of 13 drugs were used for PA development and proof of principle testing, chosen to represent major psychoactive drug classes: psychomotor stimulants (cocaine, methamphetamine, methylphenidate, mazindol, and modafinil), opioids (morphine, oxycodone, fentanyl, and codeine) and psychotomimetics (PCP, SDZ, and ketamine), with one drug that may be classified into two classes (salvinorin A, a κ -agonist opioid with psychotomimetic properties). We chose drugs of abuse as many of their psychoactive properties are well described in the clinical and preclinical literature. Within each of these drug classes we included drugs that were structurally dissimilar or known to have different pharmacological profiles. For example, methamphetamine differs from cocaine in its dopamine transporter effects (Pifl *et al*, 1995; Sandoval *et al*, 2001; Sonders *et al*, 1997; Vanderschuren *et al*, 2000) and they both differ from mazindol in their abuse potential (Chait *et al*, 1987). Morphine appears to differ from oxycodone in its prescription abuse potential (Compton and Volkow, 2006), and PCP is a noncompetitive NMDA antagonist with abuse potential whereas SDZ is a competitive antagonist with little abuse potential (Baron and Woods, 1995; Koek *et al*, 1990).

Despite this within-class diversity PA successfully discovered predictors for all three classes, thus developing a high-performance classification scheme in small groups ($n \approx 4$ animals per dose group). The classification was

generally dose dependent within each drug and drug class despite not being explicitly selected for dose effect, with the high doses scoring 3–9 standard deviations from the mean of the vehicle animals. Furthermore, in a simulation of categorizing unknown compounds PA correctly classified in a dose-dependent manner four drugs that were not used in the generation of the class predictors. In particular two problematic novel compounds, modafinil and salvinorin A, were successfully categorized according to their predominant psychopharmacological profile. Modafinil, considered a psychomotor stimulant that inhibits dopamine and norepinephrine uptake (Minzenberg and Carter, 2008), was not as fully ‘efficacious’ in its effect on our psychomotor predictor as the more typical psychomotor stimulants, and also had a very slight but still significant effect on the opioid predictor (Figure 9, left). These results could be expected given modafinil’s complex pharmacological profile and may be indicative of commonality with opioids in terms of their histaminergic effects (Ishizuka *et al*, 2008; Chikai *et al*, 1994; Barke and Hough, 1992). Salvinorin A, a κ -agonist (Roth *et al*, 2002; Chavkin *et al*, 2004) with psychotomimetic properties (Gonzalez *et al*, 2006; Dalgarno, 2007), was an interesting compound to test as it had the potential to fall into two classifications (opioid and psychotomimetic). Interestingly, PA classified it along its predominant psychopharmacological property, namely psychotomimetic (Figure 9). Only a slight opioid-like trend was detected, as opposed to drug discrimination studies demonstrating clear substitution of salvinorin A for κ -agonists (Willmore-Fordham *et al*, 2007; Li *et al*, 2008), but this hardly constitutes a failure of our current classification scheme, considering that it was not trained with any

κ -agonists before, and its opioid predictor is based predominantly on μ -agonist effects. In addition to the four test compounds a mixture of drugs (morphine and cocaine) in the same injection was tested. The outcome score was intermediate between their two classes, demonstrating that the two class predictors function at least partially as independent components. Overall, PA proved to be a general, sensitive, and specific (at the class level) *in vivo* psychopharmacological screening paradigm.

All the psychotomimetic drugs tested were found to increase turning toward the arena wall during the middle of progression segments. All opioid drugs decreased movement jerk (change in acceleration), apparently due to slow and uniform progression. All psychomotor stimulant drugs decreased slight acceleration and deceleration in low speeds, apparently due to fast crossing from 0 speed to high speed and back. It should be stressed that these class predictors depend also on the behavior in the other classes. For example, turning in all directions was equally increased by the psychotomimetics, but only turning toward the wall was isolated as a psychotomimetic predictor, because turning away from the wall was also increased by some opioids. It is thus quite possible that other, more specific predictors will be discovered when testing a wider range of drugs and classes. As with most data mining strategies, increasing the scope of the database is expected to increase the quality of the predictors that can be mined.

Interestingly it appears that using the best psychomotor stimulant predictors, low dose opioids are very similar to low dose psychomotor stimulants, while in higher doses they go their separate way (Figure 5). A similar phenomenon is seen in the cocaine–morphine mixture (Figure 9, left), in which the lower dose combination appears slightly more psychomotor-like while the higher dose combination appears more opioid-like. Separating psychomotor stimulants from opioids at low doses was the only aspect of classification in this study that PA could not be trained for, and in fact even when specifically mining for separation of these two classes only in the low dose range we could not find a single significant pattern. This suggests strong similarity of behavior at this dose range, considering that 73 042 patterns were mined. Indeed, it was shown that fentanyl can engender some discriminative stimulus properties of cocaine (Colpaert *et al*, 1979) and amphetamine produces an increasing amount of fentanyl-appropriate responding when the training doses is progressively lowered (Colpaert *et al*, 1980) in animals trained to discriminate the interoceptive cues of a drug. There is also an overlap in the net effect of psychomotor stimulants and opioids on dopamine release in the striatum (Fadda *et al*, 2005; Murphy *et al*, 2001; Zocchi *et al*, 2003). However, this overlap occurs only at doses of morphine and cocaine that would not be expected to induce detectable levels of dopamine release as measured by *in vivo* microdialysis in the C57BL/6 mouse (Rouge-Pont *et al*, 2002), although in CD1 mice the dopamine levels are comparable following low doses (Zocchi *et al*, 2003). The overlap measured in this study of the very low dose morphine (1.0 mg/kg) and fentanyl (0.056, 0.1 mg/kg) with low dose cocaine (3.0 mg/kg) may also reflect the sensitivity of PA to serotonergic mechanisms in common at these doses (George, 1990; Zocchi *et al*, 2003). Regardless, the overall sensitivity of PA

to dose, even with no explicit attempt to mine for it, illustrates the potential utility of PA as a pharmacological tool.

The problems of many behavioral animal models used for novel compound classification appear to be related to their relatively low broad-sense heritability and high interaction with confounding environmental factors. For example, Chesler *et al* (2002) estimated the heritability of the thermal nociception test, a commonly used model for assessing pain sensitivity and classify opioids, at only 27%, whereas 42% of the variance was attributed to various environmental factors (mostly the human tester) and 18% to interaction with these factors. In a study by Valdar *et al* (2006) including thousands of genetically heterogeneous mice in several common behavioral tests, all behavioral measures scored less than 41% heritability, and half of the environmental factors had interaction higher than 20%. In contrast, all the class predictors discovered in the present study were found, in a different data set of naive animals from 10 inbred strains across three laboratories, to have heritability higher than 52% and interaction with the laboratory lower than 15%. Our use of previously collected data for the heritability and replicability estimation illustrates an important advantage of the database approach—the ability to readily test any newly discovered predictors in previous data sets from various genotypes, treatments and conditions. In fact it is possible to explicitly mine behavioral patterns based on several advantageous properties as measured in different data sets, eg that they show both high-class prediction in a multidrugs data set and high heritability in a multi-genotype data set.

Several attempts have been made in the past to characterize and classify psychopharmacological substances by their effect on complex locomotor behavior. Most notably, Paulus and Geyer (1991, 1992), (see also Geyer and Paulus, 1992; Risbrough *et al*, 2006; Young *et al*, 2007) used an approach based on general considerations of dynamical systems and fractal theory to measure spatial and temporal scaling exponents of the path. There is no principle contradiction between this approach and a data mining approach, and in fact they can be complementary, eg by mining path properties in multiple scales and/or mining multiple relationships between properties in different scales (as we in part do in this study using the two attributes c_4 and c_{16} , measuring the path curvature in spatial scales of 4 and 16 cm, respectively, see Kafkafi and Elmer, 2005). The advantage of the data mining approach is that a very large number of potential variables can be explicitly screened for best prediction of any experimental factor of interest, and thus the same test can be customized and optimized to address various research questions. For example, the PA version used in this study for classifying drug-induced behavior in mice was practically identical to the PA version used for discovering early motor symptoms in *SOD1* mutant rats (a model of Amyotrophic Lateral Sclerosis, see Kafkafi *et al*, in press), except for the addition of one attribute and adapting some of the bin boundaries in the other nine attributes to their typical range in mice. Moreover, multiple research questions can in principle be investigated even in the same data set, producing a variety of different predictors for different experimental factors. For example, in the scope of the present study patterns were

not mined for any dose–response, and therefore all the dose–effect curves were produced ‘for free’, but for more advanced analysis it may be useful to explicitly mine for the strongest dose–responses by using PA with a test such as isotonic regression (Robertson *et al*, 1988) instead of *t*-test. This can be performed either within drug or within class, producing drug- or class-specific dose predictors. Subclass predictors may be mined for within a class, eg differentiating phenanthrenes (morphine, oxycodone) from anilidopiperidines (fentanyl) within opioids, or direct from indirect dopamine agonists within psychomotor stimulants. It is also possible to mine for predictors of properties across drug class, for example mining for behavioral patterns that are predictive of the drug’s abuse potential in humans. If and when they are discovered, all such predictors can easily be measured in the same session, potentially providing a detailed fingerprint of a novel compound. This ability is based on the concept that open-field behavior is complex enough to have many different aspects, and each aspect may be the most suitable for detecting a different experimental effect. Note also that, while other combinations of variables not specifically mined for class prediction may nonetheless achieve good separation of classes, the PA approach in this study also identifies specific and relatively independent predictors for each class separately (that is, the class trajectories in Figures 5, 7 and 9 are not only well separated from those of the other classes, but also generally align with the axes).

There are several methods that can be adapted for dissecting behavior into multiple types and mining them for subtle differences. An important consideration in choosing our method was the interpretability and utility of any finding. The PA method explicitly specifies the differentiating behavioral patterns, and can trace them back to particular events in the data (eg Figure 4, 6 and 8) or even in a videotaped session. It is thus possible to focus on these specific patterns, possibly reproduce them in specialized experimental setups, and study the mechanism by which they are affected. Note that in experiments focusing on few standard patterns as the endpoints of interest the analysis becomes much simpler technically, as it is not necessary anymore to measure and mine thousands of the patterns. Another important advantage of our method is that it is designed to test a broad range of patterns from the very general (single-attribute patterns, each typically occurring for about 20% or more of the total movement time) to the very specific (four-attribute patterns, each frequently occurring for less than 1%).

CNS drug discovery lags behind most areas of drug development (Kola and Landis, 2004). The difficulty in psychiatric drug development is thought to be due to several interrelated problems such as the overreliance on a few pharmacological mechanisms, the implausibility of a single causal molecular abnormality, the animals models used to screen drugs for treatment efficacy and the target-centric drug discovery approach (Spedding *et al*, 2005; Agid *et al*, 2007). The orientation of most animal models in behavioral neuroscience has been toward face and construct validity designed to provide a better understanding of the neurobiological underpinnings in the disease state. However, these types of studies are generally not ideal for the type of high-throughput screening that has

become a cornerstone technology in pharmaceutical research (Bajorath, 2002). Predictive animal models better fit the requirements of a high-throughput environment, but a drawback to most types of these models is that they are restricted to the identification of drugs that antagonize either a narrow pharmacology, block a narrow behavior or show utility for a single drug class. Novel compound identification might be severely limited under these conditions. *In vitro* assays have been successfully used as predictive models in toxicology and cancer response (Golub *et al*, 1999; Thomas *et al*, 2007), and a critical catalyst in their success has been the incorporation of data mining strategies. This type of strategy has recently been used in a human neuronal precursor cell line as a means to discover gene-expression patterns predictive of a drug’s psychoactive class. In these studies (Gunther *et al*, 2003, 2005) various classification algorithms were used to discover gene-expression profiles that were predictive of antidepressant, antipsychotic, and opioid drug action. Although this is a significant step forward, an *in vitro* system is necessarily restricted in its ability to mimic a true system pharmacology response to psychiatric drugs, and would ideally be complemented with an *in vivo* assay for more advanced preclinical testing. Our *in vivo* data mining approach establishes a framework for screening novel chemical entities based on a response system that represents the net pharmacological effect on the system of interest, namely the CNS.

ACKNOWLEDGEMENTS

This study was supported by NIH Cutting Edge Basic Research Award (CEBRA grant DA022407).

DISCLOSURE/CONFLICT OF INTEREST

The author(s) declare that, except for income received from the primary employer, no financial support or compensation has been received from any individual or corporate entity over the past 3 years for research or professional service and there are no personal financial holdings that could be perceived as constituting a potential conflict of interest. Dr Elmer and Dr Kafkafi have applied for a patent related to the Pattern Array procedure described in the paper (US Patent Application no. 11/634,068 (publication no. US 2007/0265816), ‘System and Method for Identifying Behavioral Signatures’). Dr Yekutieli has no conflicts to disclose.

REFERENCES

- Agid Y, Buzsáki G, Diamond DM, Frackowiak R, Giedd J, Girault J *et al* (2007). How can drug discovery for psychiatric disorders be improved? *Nat Rev Drug Discov* 6: 189–201.
- Arnt J (1982). Pharmacological specificity of conditioned avoidance response inhibition in rats: inhibition by neuroleptics and correlation to dopamine receptor blockade. *Acta Pharmacol Toxicol (Copenh)* 51: 321–329.
- Bajorath J (2002). Integration of virtual and high-throughput screening. *Nat Rev Drug Discov* 1: 882–894.

- Barke KE, Hough LB (1992). Morphine-induced increases of extracellular histamine levels in the periaqueductal grey *in vivo*: a microdialysis study. *Brain Res* 572: 146–153.
- Baron SP, Woods JH (1995). Competitive and uncompetitive *N*-methyl-D-aspartate antagonist discriminations in pigeons: CGS 19755 and phencyclidine. *Psychopharmacology (Berl)* 118: 42–51.
- Benjamini Y, Yekutieli D (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J Am Stat Assoc* 100: 71.
- Broadhurst PL (1975). The Maudsley reactive and nonreactive strains of rats: a survey. *Behav Genet* 5: 299–319.
- Brunner D, Nestler E, Leahy E (2002). In need of high-throughput behavioral systems. *Drug Discov Today* 7(18 Suppl): S107–S112.
- Butcher EC, Berg EL, Kunkel EJ (2004). Systems biology in drug discovery. *Nat Biotechnol* 22: 1253–1259.
- Chait LD, Uhlenhuth EH, Johanson CE (1987). Reinforcing and subjective effects of several anorectics in normal human volunteers. *J Pharmacol Exp Ther* 242: 777–783.
- Chavkin C, Sud S, Jin W, Stewart J, Zjawiony JK, Siebert DJ et al (2004). Salvinorin A, an active component of the hallucinogenic sage *Salvia divinorum* is a highly efficacious kappa-opioid receptor agonist: structural and functional considerations. *J Pharmacol Exp Ther* 308: 1197–1203.
- Chesler EJ, Wilson SG, Lariviere WR, Rodriguez-Zas SL, Mogil JS (2002). Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neurosci Biobehav Rev* 26: 907–923.
- Chikai T, Oishi R, Saeki K (1994). Increase in the extracellular histamine concentration in the rat striatum by mu-opioid receptor activation. *J Neurochem* 62: 724–729.
- Clarke PA, te Poele R, Workman P (2004). Gene expression microarray technologies in the development of new therapeutic agents. *Eur J Cancer* 40: 2560–2591.
- Colpaert FC, Niemegeers CJ, Janssen PA (1979). Discriminative stimulus properties of cocaine: neuropharmacological characteristics as derived from stimulus generalization experiments. *Pharmacol Biochem Behav* 10: 535–546.
- Colpaert FC, Niemegeers CJ, Janssen PA (1980). Factors regulating drug cue sensitivity: the effect of training dose in fentanyl-saline discrimination. *Neuropharmacology* 19: 705–713.
- Compton WM, Volkow ND (2006). Major increases in opioid analgesic abuse in the United States: concerns and strategies. *Drug Alcohol Depend* 81: 103–107.
- Creese I, Burt DR, Snyder SH (1976). Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. *Science* 192: 481–483.
- Dalgarno P (2007). Subjective effects of *Salvia divinorum*. *J Psychoactive Drugs* 39: 143–149.
- Drai D, Golani I (2001). SEE: a tool for the visualization and analysis of rodent exploratory behavior. *Neurosci Biobehav Rev* 25: 409–426.
- Fadda P, Scherma M, Fresu A, Collu M, Fratta W (2005). Dopamine and serotonin release in dorsal striatum and nucleus accumbens is differentially modulated by morphine in DBA/2J and C57BL/6J mice. *Synapse* 56: 29–38.
- George FR (1990). Cocaine produces low dose locomotor depressant effects in NBR and F344 rats. *Pharmacol Biochem Behav* 37: 795–798.
- Geyer MA, Paulus MP (1992). 'Multivariate and nonlinear approaches to characterizing drug effects on the locomotor and investigatory behavior of rats'. *NIDA Res Monogr* 124: 203–235.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP et al (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531–537.
- Gonzalez D, Riba J, Bouso JC, Gomez-Jarabo G, Barbanj MJ (2006). Pattern of use and subjective effects of *Salvia divinorum* among recreational users. *Drug Alcohol Depend* 85: 157–162.
- Gunther EC, Stone DJ, Gerwien RW, Bento P, Melvyn PH (2003). Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles *in vitro*. *Proc Natl Acad Sci USA* 100: 9608–9613.
- Gunther EC, Stone DJ, Rothberg JM, Gerwien RW (2005). A quantitative genomic expression analysis platform for multiplexed *in vitro* prediction of drug action. *Pharmacogenomics J* 5: 126–134.
- Hen I, Sakov A, Kafkafi N, Golani I, Benjamini Y (2004). The dynamics of spatial behavior: how can robust smoothing techniques help? *J Neurosci Methods* 133: 161–172.
- Hood L, Perlmutter RM (2004). The impact of systems approaches on biological problems in drug discovery. *Nat Biotechnol* 22: 1215–1217.
- Horev G, Benjamini Y, Sakov A, Golani I (2007). Estimating wall guidance and attraction in mouse free locomotor behavior. *Gene Brain Behav* 6: 30–41.
- Ishizuka T, Murakami M, Yamatodani A (2008). Involvement of central histaminergic systems in modafinil-induced but not methylphenidate-induced increases in locomotor activity in rats. *Eur J Pharmacol* 578: 209–215.
- Kafkafi N, Benjamini Y, Sakov A, Elmer GI, Golani I (2005). Genotype–environment interactions in mouse behavior: a way out of the problem. *Proc Natl Acad Sci USA* 102: 4619–4624.
- Kafkafi N, Elmer GI (2005). Texture of locomotor path: a replicable characterization of a complex behavioral phenotype. *Genes Brain Behav* 4: 431–443.
- Kafkafi N, Lipkind D, Benjamini Y, Mayo CL, Elmer GI, Golani I (2003a). SEE locomotor behavior test discriminates C57BL/6J and DBA/2J mouse inbred strains across laboratories and protocol conditions. *Behav Neurosci* 117: 464–477.
- Kafkafi N, Mayo C, Drai D, Golani I, Elmer G (2001). Natural segmentation of the locomotor behavior of drug-induced rats in a photobeam cage. *J Neurosci Meth* 109: 111–121.
- Kafkafi N, Pagis M, Lipkind D, Mayo CL, Benjamini Y, Elmer GI et al (2003b). Darting behavior: a quantitative movement pattern for discrimination and replicability in mouse locomotor behavior. *Behav Brain Res* 142: 193–205.
- Kafkafi N, Yekutieli D, Yarowsky Y, Elmer GI. Data mining in a behavioral test detects early symptoms in a model of amyotrophic lateral sclerosis. *Behav Neurosci* (in press).
- Koek W, Woods JH, Colpaert FC (1990). *N*-methyl-D-aspartate antagonism and phencyclidine-like activity: a drug discrimination analysis. *J Pharmacol Exp Ther* 253: 1017–1025.
- Kola I, Landis J (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3: 711–715.
- Kuribara H, Tadokoro S (1981). Correlation between antiavoidance activities of antipsychotic drugs in rats and daily clinical doses. *Pharmacol Biochem Behav* 14: 181–192.
- Minzenberg MJ, Carter CS (2008). Modafinil: a review of neurochemical actions and effects on cognition. *Neuropsychopharmacology* 33: 1477–1502.
- Murphy NP, Lam HA, Maidment NT (2001). A comparison of morphine-induced locomotor activity and mesolimbic dopamine release in C57BL6, 129Sv and DBA2 mice. *J Neurochem* 79: 626–635.
- Leppanen PK, Ewalds-Kvist SB, Selander RK (2005). Mice selectively bred for open-field thigmotaxis: life span and stability of the selection trait. *J Gen Psychol* 132: 187–204.
- Li JX, Rice KC, France CP (2008). Discriminative stimulus effects of 1-(2,5-dimethoxy-4-methylphenyl)-2-aminopropane in rhesus monkeys. *J Pharmacol Exp Ther* 324: 827–833.
- Lipkind D, Sakov A, Kafkafi N, Elmer GI, Benjamini Y, Golani I (2004). New replicable anxiety-related measures of wall vs center behavior of mice in the open field. *J Appl Physiol* 97: 347–359.

- Paulus MP, Geyer MA (1991). A temporal and spatial scaling hypothesis for the behavioral effect of psychostimulants. *Psychopharmacology* **104**: 6–16.
- Paulus MP, Geyer MA (1992). ‘The effects of MDMA and other methylenedioxy-substituted phenylalkylamines on the structure of rat locomotor activity’. *Neuropsychopharmacology* **7**: 15–31.
- Pifl C, Drobný H, Reither H, Hornykiewicz O, Singer EA (1995). Mechanism of the dopamine-releasing actions of amphetamine and cocaine: plasmalemmal dopamine transporter versus vesicular monoamine transporter. *Mol Pharmacol* **47**: 368–373.
- Risbrough VB, Masten VL, Caldwell S, Paulus MP, Low MJ, Geyer MA (2006). Differential contributions of dopamine D1, D2, and D3 receptors to MDMA-induced effects on locomotor behavior patterns in mice. *Neuropsychopharmacology* **31**: 2349–2358.
- Robertson T, Wright FT, Dykstra RL (1988). *Order restricted statistical inference*. Wiley series in probability and mathematical statistics. Wiley: New York.
- Roth BL, Baner K, Westkaemper R, Siebert D, Rice KC, Steinberg S et al (2002). A potent naturally occurring nonnitrogenous kappa opioid selective agonist. *Proc Natl Acad Sci USA* **99**: 11934–11939.
- Rouge-Pont F, Usiello A, Benoit-Marand M, Gonon F, Piazza PV, Borrelli E (2002). Changes in extracellular dopamine induced by morphine and cocaine: crucial control by D2 receptors. *J Neurosci* **22**: 3293–3301.
- Sandoval V, Riddle EL, Ugarte YV, Hanson GR, Fleckenstein AE (2001). Methamphetamine-induced rapid and reversible changes in dopamine transporter function: an *in vitro* model. *J Neurosci* **21**: 1413–1419.
- Sonders MS, Zhu SJ, Zahniser NR, Kavanaugh MP, Amara SG (1997). Multiple ionic conductances of the human dopamine transporter: the actions of dopamine and psychostimulants. *J Neurosci* **17**: 960–974.
- Spedding M, Jay T, Costa e Silva J, Perret L (2005). A pathophysiological paradigm for the therapy of psychiatric disease. *Nat Rev Drug Discov* **4**: 467–476.
- Tan PN, Steinbach M, Kumar V (2006). *Introduction to Data Mining*. Classification: basic concepts, decision trees and model evaluation. Addison-Wesley: Boston.
- Thomas RS, O’Connell TM, Pluta L, Wolfinger RD, Yang L, Page TJ (2007). A comparison of transcriptomic and metabonomic technologies for identifying biomarkers predictive of two-year rodent cancer bioassays. *Toxicol Sci* **96**: 40–46.
- Tecott LH, Nestler EJ (2004). Neurobehavioral assessment in the information age. *Nat Neurosci* **7**: 462–466.
- van der Greef J, McBurney RN (2005). Innovation: rescuing drug discovery: *in vivo* systems pathology and systems pharmacology. *Nat Rev Drug Discov* **4**: 961–967.
- Vanderschuren LJ, Schoffelmeer AN, Wardeh G, De Vries TJ (2000). Dissociable effects of the kappa-opioid receptor agonists bremazocine, U69593, and U50488H on locomotor activity and long-term behavioral sensitization induced by amphetamine and cocaine. *Psychopharmacology (Berl)* **150**: 35–44.
- Valdar W, Solberg LC, Gauguier D, Cookson WO, Rawlins JN, Mott R et al (2006). Genetic and environmental effects on complex traits in mice. *Genetics* **174**: 959–984.
- Wadenberg ML, Hicks PB (1999). The conditioned avoidance response test re-evaluated: is it a sensitive test for the detection of potentially atypical antipsychotics? *Neurosci Biobehav Rev* **23**: 851–862.
- Willmore-Fordham CB, Krall DM, McCurdy CR, Kinder DH (2007). The hallucinogen derived from *Salvia divinorum*, salvinorin A, has kappa-opioid agonist discriminative stimulus effects in rats. *Neuropharmacology* **53**: 481–486.
- Willner P (1991). Methods for assessing the validity of animal models of human psychopathology. In: Boulton AA, Baker GB, Martin-Iversen MT (eds). *Animal Models in Psychiatry, I*. 18th edn. Humana Press: New Jersey. pp 1–24.
- Yen-Koo HC, Davis DA, Balazs T (1985). Inhibition of dopaminergic agonist-induced gnawing behavior by neuroleptic drugs in mice. *Drug Chem Toxicol* **8**: 495–502.
- Young JW, Minassian A, Paulus MP, Geyer MA, Perry W (2007). A reverse-translational approach to bipolar disorder: rodent and human studies in the Behavioral Pattern Monitor. *Neurosci Biobehav Rev* **31**: 882–896.
- Zocchi A, Girlanda E, Varnier G, Sartori I, Zanetti L, Wildish GA et al (2003). Dopamine responsiveness to drugs of abuse: a shell-core investigation in the nucleus accumbens of the mouse. *Synapse* **50**: 293–302.