`

Beyond locomotion: in the mouse the mapping between sensations and behaviours unfolds in a higher dimensional space

Riccardo Storchi[1,2,*], Timothy F. Cootes[1,3], Robert J. Lucas[1,2]

[1] University of Manchester, Faculty of Biology, Medicine and Health

[2] School of Biological Science, Division of Neuroscience & Experimental Psychology

[3] School of Health Science, Division of Informatics, Imaging & Data Science

[*]Corresponding Author: riccardo.storchi@manchester.ac.uk

Abstract

The ability of specific sensory stimuli to evoke stereotyped behavioural responses in the mouse represents an attractive opportunity to study how the brain drives behaviour. To achieve its potential, this approach relies not only upon advanced methods to measure and manipulate neurophysiological activity, but also upon a complete and unbiased description of the sensory evoked actions. In practice measuring motor actions in a freely moving mouse has proven technologically challenging and responses such as escape or freeze have been simply defined by measuring locomotion. Here we develop a method to quantify a wider array of motor actions and to ask whether this higher dimensional representation of behavioral responses captures additional aspects of auditory and visually guided behaviours. We find that locomotion alone is not sufficient to capture the sensory specificity of behavioural responses and a higher dimensional representation of postures and movements is required to map between sensations and behaviours.

`

Introduction

The study of sensory guided actions represents a powerful approach to understanding the neurophysiological basis of behavior. Such spontaneous behaviors as exploration (Cooke et al., 2015; Smeds et al., 2019), hunting (Hoy et al., 2016; Hoy et al., 2019), and escape and freeze (De Franceschi et al., 2016; Evans et al., 2018; Storchi et al., 2019; Yilmaz and Meister, 2013), while heterogeneous, share the key property that they can be reproducibly elicited in the lab by controlled sensory stimulation. The ability of sensory stimuli to evoke a reproducible behavioural response in these paradigms makes them an important experimental tool to understand how inputs are encoded and interpreted in the brain, and appropriate actions selected.

Realizing the full power of this approach, however, relies upon a description of evoked behaviors that is sufficiently complete to encompass the full complexity of the motor responses and to capture the relevant variations across different stimuli or repeated presentations of the same stimulus. Behaviours such as sensory guided escape or freeze have been defined on the basis of a clear phenotype – a sudden change in locomotion state. Thus in the last few years it has been shown that speed, size, luminance and contrast of a looming object have different and predictable effects on locomotion (Evans et al., 2018; Storchi et al., 2019; Yilmaz and Meister, 2013). Nevertheless, mice do more than run, and a variety of other body movements as well as changes in body orientation and posture could, at least in principle, contribute to sensory evoked behaviours. Until now, a lack of tools to objectively measure types of movement other than locomotion has left that possibility unexplored. We set out here to ask whether a richer representation of mouse behaviour was possible and, if so, whether this provided additional information about the relationship between sensation and actions. To this end we developed a system based on the 3D reconstruction of mouse poses. We then applied this system to generate a higher dimensional representation of mouse sensory guided behaviours that enables the quantification of a wide range of body movements and postures.

We found that behavioural responses to simple visual and auditory stimuli encompass numerous motor actions. Since different motor actions are not always correlated, accounting for all those actions provides a richer description of behaviour by increasing the dimensionality of behavioural representation. This increase in dimensionality provides an improved understanding of sensory guided behaviours in several respects. First, behavioural responses are more specific to distinct stimuli than is apparent simply by measuring locomotion. Second, this increased specificity can be explained at least in part by the appearance of a richer repertoire of behaviours, with equivalent locomotor responses found to differ in other behavioural dimensions. Third, each class of sensory stimulus can evoke more than one type of behaviour, revealing a ‘one to many’ map between stimulus and response.

`

Results

A method for quantifying motor actions

The first aim of this study was to develop a method to quantify a wide range of movements and postures. Five different landmarks on the mouse body (nose tip, left & right ears, neck base and tail base) were tracked by using four cameras (**Fig. 1a**, **Supplementary Fig. 1a,b**). The 3D pose of the animal was then reconstructed by using a method we established for this study (**Fig. 1b**; **Supplementary Fig. 1c-f; Supplementary Movie 1;** see **Methods**). These pre-processing stages allowed us to describe, on a frame-by-frame basis, the mouse pose $\boldsymbol{X}$ as

$$\boldsymbol{X}(t) = \left(\overline{\boldsymbol{X}} + \sum_{i=1}^{N_{eigenposes}} \boldsymbol{P}_i b_i(t) \right) \boldsymbol{R}(t) + \boldsymbol{T}(t) \quad (1)$$

Where: t represents the time of the current frame; $\boldsymbol{X}$ the coordinates of the body landmarks; $\overline{\boldsymbol{X}}$ the body coordinates of a reference pose; b_i the shape parameters allowing to keep track of the changes in the body shape (**Fig. 1c**); $\boldsymbol{R}$ and $\boldsymbol{T}$ the rigid transformations (rotation and translation) encoding the animal's position in the behavioural arena (**Fig. 1c**). Both $\overline{\boldsymbol{X}}$ and $\boldsymbol{P}_i$ (the eigenposes) were obtained by training a Statistical Shape Model on a separate dataset of mouse poses (see **Methods**). The first two eigenposes captured respectively body elongation and bending (**Fig. 1c**), two important descriptors of the mouse posture.

Based on this analytical description of the mouse pose we then developed two sets of measures to quantify distinct postures and movements. The first set of measures, rearing, body elongation and body bending, allowed us to capture different aspects of the mouse posture (**Fig. 1d, Postural Measures**). The second set, constituted by locomotion, freezing, rigid body rotation and changes in rearing, body elongation and body bending allowed us to capture different types of body movements (**Fig. 1d, Movement Measures**).

Measuring multiple motor actions allows a higher dimensional representation of behaviour

We then set out to investigate the extent to which those body postures and movements were involved in sensory guided behaviours. The animals were tested in an open field arena in which no shelter was provided. In order to capture a wide range of behavioural responses we used three different classes of sensory stimuli, out of which two were visual and one auditory. Among visual stimuli we selected a bright flash and a looming object. We have previously shown that these two stimuli evoked distinct and opposite behavioural response, with the former inducing an increase in locomotor activity while the latter abolishes locomotion by inducing freezing behaviour (Storchi et al., 2019). The auditory stimulus was also previously shown to induce defensive responses such as freeze or startle (Evans et al., 2018; Koch, 1999).

We separately averaged all trials according to stimulus class and we found that all our measures of body postures and movements were involved in sensory guided behaviours (**Fig. 2a**). Across most measures average response to the flash clearly diverged from those elicited

`

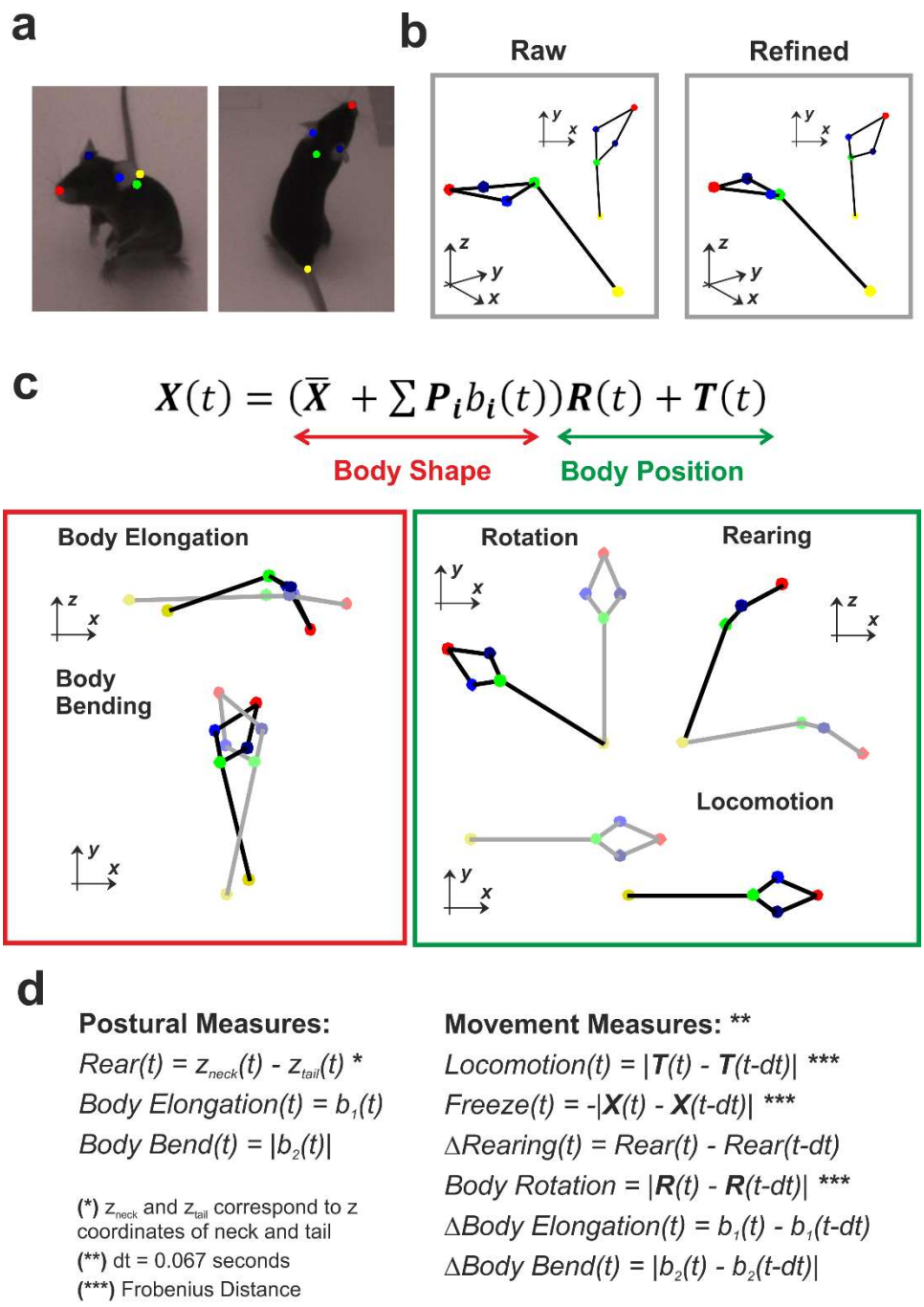

Figure 1: Description of the model used to quantify mouse behaviour. a) Body landmarks are separately tracked across each camera. **b)** A raw 3D reconstruction is obtained by triangulation of body landmark positions (left panel). The raw reconstruction is corrected by applying our algorithm based on the Statistical Shape Model as described in Methods. The refined 3D reconstruction (right panel) is then used for all the further analyses. **c)** The model expressed by equation 1 allows for quantifying a wide range of postures and movements of which red and green boxes report some examples. The "Body Shape" components enable to measure changes in body shape such as body elongation and body bending. The "Body Position" components enables to quantify translations and rotations in a 3D space. **d)** The full set of behavioural measures, divided into 3 postural measures and 6 movement measures is expressed as function of the terms in equation 1.

`

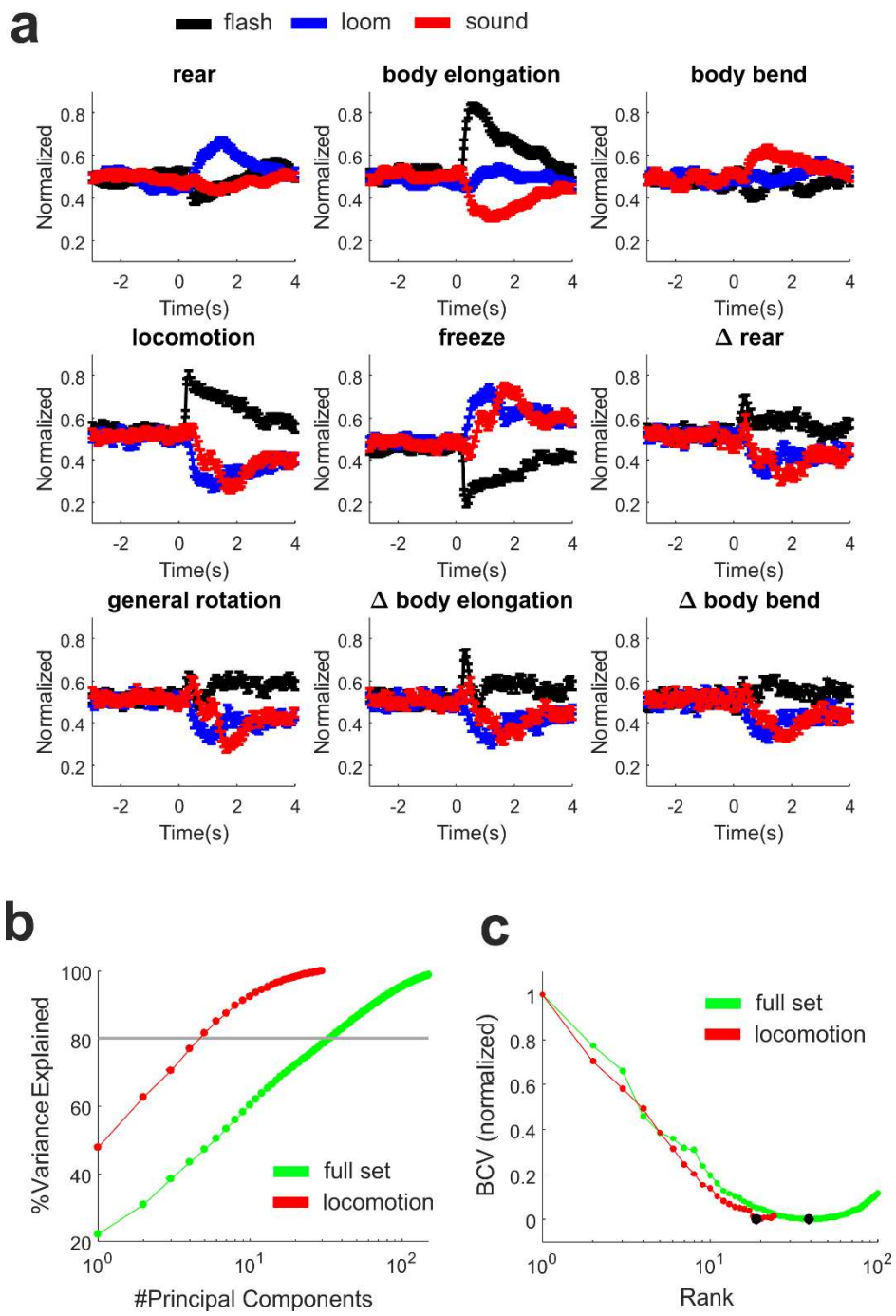

Figure 2: **Multiple motor actions are involved in sensory guided behaviours a)** Average response to the three classes of sensory stimuli (flash, loom, sound). Error bars represent SEM (n=172 for each stimulus class). **b)** Percentage of variance explained as function of principal components for the full set of motor actions (green) and for locomotion only (red). For the full set 34 components are required to explain >80% variance (gray horizontal line), for locomotion only 5 components. **c)** The minimum Bi-Cross Validation Error is used to quantify the rank of the full set and of locomotion only (respectively 19 and 39, marked by black dots). Both principal components and Bi-Cross Validation are performed on the epoch 0-2 seconds from stimulus onset.

by other stimuli. Average responses to looming and sound were instead similar for some measures (e.g. locomotion, freeze) and more distinct for others (e.g. rear, body elongation).

`

The appearance of distinct, imperfectly correlated, motor actions raises the possibility that inclusion of all our measures of movements and postures, hereafter the full set, may increase the dimensionality of our behavioural description. To determine whether this was the case, we performed a Principal Component Analysis (PCA) and compared the number of dimensions required to explain the variance in behavioural responses pooled across stimuli when considering all our measures vs locomotion alone (**Fig. 2b**). This analysis revealed that, in order to explain >80% variance in the full set, 34 dimensions were required while 5 dimensions were sufficient for locomotion alone. In principle the increase in dimensionality observed in the full set could be trivially explained by a disproportionate increase in measurement noise. To test for this possibility, we estimated the rank of the response matrix by applying the Bi-Cross Validation technique (Owen and Perry, 2009). Consistent with the PCA analysis, we found that the rank of the full set was substantially larger, ~two-fold (**Fig. 2c**), indicating that the full set provided a genuine increase in dimensionality.

Higher dimensionality reveals increased specificity in sensory guided behaviours

We then asked whether this increased dimensionality could capture additional aspects of stimulus-response specificity that could not be observed in locomotion. To account for the fact that evoked responses developed over time for all the following analysis we divided the responses into three consecutive epochs of 1s duration according to their latency from the stimulus onset.

We first looked for a specific condition in which the same level of locomotion was expressed in response to two distinct sensory stimuli. A simple illustrative example where locomotion largely fails to capture stimulus-response specificity is the case in which both looming and sound induce a common freezing pattern (**Fig. 3a**, left panel). In the intermediate epoch (1-2s after stimulus onset), when freezing is strongest, locomotion "saturates" towards 0 in responses to both stimuli, and thus provides no discrimination. However, stimulus-specificity is apparent in the animal's posture as revealed by quantifying body elongation (**Fig. 3a**, right panel; see also **Supplementary Movies 2, 3**).

In order to systematically compare stimulus-response specificity for the full set of postures and movements with that revealed by locomotion alone we developed a simple Specificity Index (SI, **Fig. 3b**). On an individual trial basis SI identified the most similar behavioural responses across our dataset and quantified the fraction of those responses that were associated with the same stimulus. The SI was calculated for pairwise comparisons between the 3 sensory stimuli. Thus, on a given trial and for each pair of stimuli, **SI** ranged from 0 to 1, in which 1 signifies all similar behavioural responses being elicited by the same stimulus, 0.5 similar responses being expressed for both stimuli, and 0 when similar responses are always associated with the other stimulus (**Fig. 3b**). We compared the distribution of SI values obtained when behaviour was described in terms of the full set of measures of postures and movements, with the SI values obtained by using only locomotion. We found that SI was always comparable or higher when calculated on the full set (**Fig. 3c**).

`

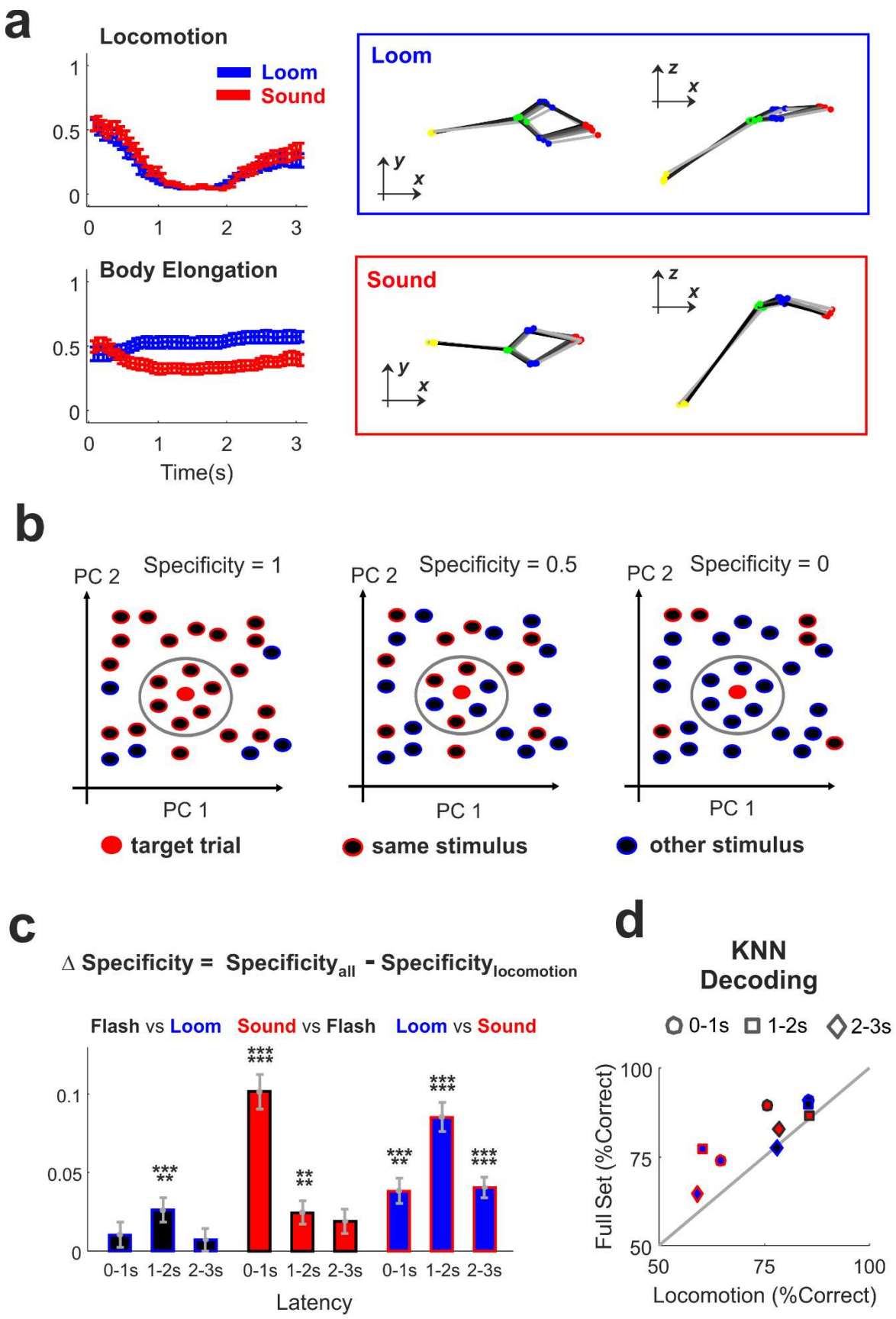

Figure 3: Higher dimensionality reveals a larger set of sensory specific behaviours. a) Loom and sound could evoke an indistinguishable pattern of locomotion arrest shown in upper left panel (mean±sem; n = 37, 31 trials, respectively loom and sound). However the pattern of body elongation

`

was different across loom and sound (bottom left panel; mean±sem). A representative trial for loom (blue box) and for sound (red box) are reported in the right panels. Time progression is captured by the gray-to-black transition of the mouse body (poses sampled every 0.2s between 1 and 2 seconds latency from stimulus onset). Note that different levels of body elongation can be observed from a side view in the z-x planes. **b)** On each trial the specificity index was calculated as the number of neighbour responses to the same stimulus divided by the total number of neighbours. In this toy example, based on two-dimensional responses (PC1 and PC2), we show a target trial for which the number of neighbouring responses for the same stimulus changes across panels to obtain Specificity Index values of 1, 0.5 and 0. **c)** For each trial the Specificity Index for locomotion ($Specificity_{locomotion}$, for k = 30 neighbours) was subtracted from the value obtained for the full set of measures of postures and movements ($Specificity_{all}$) to obtain ΔSpecificity. This measure was significantly positive across a number of conditions (Flash vs Loom: $p < 0.0005$ at latency 1-2s; Sound vs Flash: $p < 0.0001$ and $p < 0.001$ at latency 0-1s and 1-2s; Loom vs Sound: $p < 0.0005$, $p < 0.0001$ and $p < 0.0001$ at latency 0-1s, 1-2s and 2-3s) and otherwise not different from zero. **d)** Comparison between K-Nearest Neighbour (KNN) decoding based on the full set and on locomotion only. Decoding accuracy is expressed as percent of correct predictions and results based on 10 fold cross-validation (repeated 100 times). Pairwise comparisons for flash vs loom (black-blue), sound vs flash (red-black) and loom vs sound (blue-red). Circles, squares and diamonds represent different epochs of the behavioural response (0-1s, 1-2s, 2-3s from stimulus onset). KNN decoding, when applied to the full set of motor actions and postures, provided substantially better decoding results (on average 9.55±8.99sd%; $p = 0.039$, signtest for n = 9 comparisons) compared with the decoding results obtained by using only locomotion.

As a further test of the conclusion that a higher dimensional description of behaviour revealed increased stimulus-response specificity, we asked whether it improved our ability to infer the stimulus class based upon our readout of behaviour – i.e. whether higher dimensionality enables more accurate decoding of the stimulus. To this end, we applied a decoding algorithm that could readily utilize the local information that was present in the high-dimensional space of behavioural responses. Since the Specificity Index constitutes the input for a K-Nearest Neighbours (KNN) classifier we adopted this algorithm as a natural extension of our specificity analysis. By systematically varying the number of neighbours we compared the best performances obtained for the full set of measures with that achieved using locomotion alone. We found that the full set of measures allowed for a substantial increase in decoding accuracy (**Fig. 3d**). In order to assess the robustness of this result we also tested two other decoding algorithms: individual Classification Trees and Random Forests. Both KNN and Random Forests outperformed Classification Trees (**Supplementary Fig. 2a,b**). The KNN decoder performed similarly to the Random Forests (**Supplementary Figure 2c**) indicating that a simple non-parametric approach based on local information represents an effective strategy to extract the sensory specific information from our dataset. Like KNN, Random Forest decoding revealed that the full set of measures systematically provided better decoding compared with locomotion alone (**Supplementary Fig. 2d**).

Higher dimensionality reveals a larger set of sensory specific behaviours

These results indicate that the mapping between stimulus and behavioural response is more specific in a higher dimensional space. We next sought to describe the structure of this mapping. Specifically, we asked how many distinct behaviours are expressed in response to each class of sensory stimulus. For this purpose, we started by clustering responses from all

`

trials based upon similarity in motor actions. An important consideration in such a process is how many clusters to allow. We approached that problem by investigating the relationship between the number of clusters and the degree to which each cluster was restricted to a single stimulus (quantified as Mutual Information between stimulus and behavioural response). We focussed on the interval 0-2s since these latencies provided the best decoding results. To increase the robustness of this analysis a K-means++ algorithm was repeated 50 times for each cluster size. Then, for each number of clusters, we estimated the Mutual Information (MI) between stimulus and behavioural response. By observing the increase in MI as function of number of clusters two distinct regions could be clearly delineated (**Fig. 4a**). For a small number of clusters, approximately between 2 and 7, we observed a "high gain" region where MI increases substantially for each additional cluster. Beyond this domain the "high gain" region was replaced by a "low gain" region where further increments in the number of clusters provided limited increments in MI. This analysis suggests 7 clusters as a reasonable trade-off between the need for a generalization of the behavioural responses and the granularity required to capture a large fraction of stimulus specific information.

Our previous analyses strongly suggest that the range of behaviours is larger when considering the full set of motor actions than simply locomotion alone. To confirm that this was true, we applied the same clustering method to the locomotion data alone. A similar repartition into high and low gain regions was observed (**Fig. 4b**). However, the high gain region domain appeared to be reduced to approximately 3-4 clusters suggesting a reduction in the number of sensory specific behavioural clusters. To more rigorously test whether this was the case we fitted MI with a function that incorporates a steep exponential component and a more gradual linear component to account respectively for the high and the low domain regions (see **Methods**). We then used the exponential rise constant $\boldsymbol{\tau}$ as a measure of the size of the high domain region. We found that $\boldsymbol{\tau}$ was indeed smaller for locomotion alone (**Fig. 4c**) indicating that the full set of measures of postures and movements captures a larger number of sensory specific behaviours.

Among the 7 behaviours revealed by our clustering of the full set several motifs occurred (**Fig. 4d**). Fast sustained locomotion (cluster #1) or rearing (cluster #2) both accompanied by body elongation; Body bending followed by delayed freeze (cluster #3); Sustained freeze (cluster #4); Transient freeze in rearing position (cluster #5); Body bending and other rotations of the body axis, including frequent changes in rearing position (cluster #6); Sustained freeze in body bent positions (cluster #7).

The Flash stimulus evoked behaviours that were very specific for this stimulus (cluster #1 and #2; **Fig. 4e**, left panel). The Loom and Sound stimuli evoked the approximately the same set of behaviours but, between the two stimulus classes, those behaviours were expressed in different proportions (**Fig. 4e**, middle and left panel).

All together our results show that a more detailed quantification of postures and movements captures aspects of sensory-guided behaviours that are distinct from changes in locomotion. Including these additional aspects reveals that stimulus specificity is higher than previously thought.

`

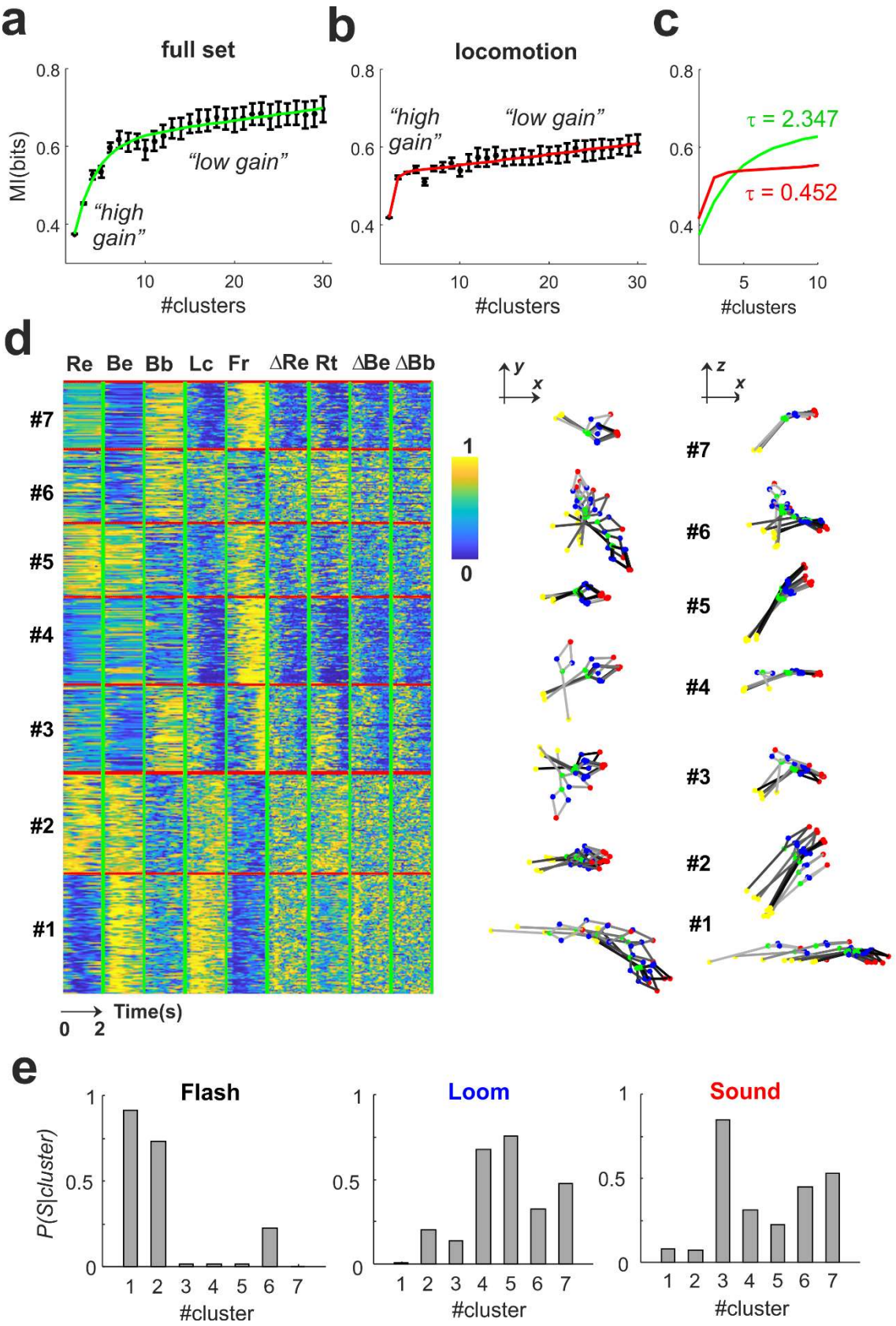

Figure 4: Higher dimensionality reveals a larger set of sensory specific behaviours. a) Mutual information is estimated for the full set of motor actions as function of the number of clusters obtained with the k-means++ algorithm (mean±sd, 50 repeats per cluster, at each repeat the best of

`

100 runs was selected). Note an initial fast rise in MI ("high gain" region in the plot) followed by a more gradual linear increase ("low gain" region). Based on these results the relation between MI and the number of clusters k was fit by using the function $MI(k) = a\left(1 - e^{-\frac{k}{\tau}}\right) + bx$. The parameter τ captures the exponential rise in Mutual Information in the "high gain" domain. **b)** Same as panel (a) but for locomotion only. **c)** Comparison between the exponential rise in MI for the full set of motor actions and for locomotion only. The exponential rise in MI, captured by the τ values, is slower for the full set indicating that the high gain domain encompasses a larger number of distinct clusters. **d)** Left panel shows the response matrix of the full dataset (n=516 trials) partitioned into 7 clusters. The response matrix is obtained by concatenating all the postures and motor actions (Re = Rear; Be = Body elongation; Bb = Body bend; Lc = Locomotion; Fr = Freeze; ΔRe = ΔRear; Rt = Body rotation; ΔBe = ΔBody elongation; ΔBb = ΔBody bend). Right panels shows one representative trial for each cluster (10 poses sampled at 0.2s intervals between 0 and 2s latency from stimulus onset; time progression is captured by the gray-to-black transition). **e)** Conditional probability of stimulus class, given each of the clusters shown in panel (d). Flash, Loom and Sound are reported respectively in left, middle and right panel.

`

Discussion

A fundamental goal of neuroscience is to link neural circuits to behaviours. Two unescapable tasks are essential prerequisites for approaching this problem: the generation of a detailed anatomical and physiological description of brain circuits – the neural repertoire – and the charting of all the relevant behaviours exhibited by the model organism of choice – the behavioural repertoire. Then, in order to uncover meaningful links, the resolutions of the neural and the behaviour repertoires have to match, since a high resolution on one side can't compensate for low resolution on the other (Krakauer et al., 2017).

In the last decade enormous advances have been made in understanding functional and anatomical connectivity of the CNS (Bernstein and Boyden, 2011). By contrast, the behavioural repertoire of a key model organism such as the mouse is still low resolution and urgently need to catch up. A powerful approach to tackle the complexity of the behavioural repertoire is to study sensory guided behaviours. Those behaviours are spontaneous, ethologically relevant and can be elicited in a lab within a delimited time window defined by the experimenter through a controlled manipulation of the sensory environment. A detailed sketching of the neural repertoire underlying these behaviours is in process and substantial advances have been made in the last few years (Evans et al., 2018; Reinhard et al., 2019; Shang et al., 2018; Shang et al., 2015; Tovote et al., 2016; Wei et al., 2015). However the behavioural repertoire is still coarse and all studies so far relied on a single behavioural phenotype constituted by a change in the locomotion state.

The first aim of this work was to provide a higher resolution map of sensory guided behaviours by developing a simple 3D model of the mouse body. The model we propose can be easily fit to the behavioural data and provides a set of clearly interpretable measures of mouse postures and movements. The set of behavioural measures we used, ranging from body elongation, bending and rearing to more commonly employed locomotion and freezing provide a higher dimensional description of mouse behaviour. In the long run this approach could allow for collecting more information from individual animals and potentially for reducing the number of animals required for preclinical testing. In this study we employed this model to quantify sensory guided responses to relevant visual and auditory stimuli that have been largely employed both in basic science (Yilmaz and Meister, 2013; Cooke et al., 2015; De Franceschi et al., 2016; Hoy et al., 2016; Evans et al., 2018; Smeds et al., 2019; Hoy et al., 2019) and for preclinical studies (Davis et al., 1993; Storchi et al., 2019).

Our first result was that all those behavioural measures are to some extent stimulus specific, i.e. they differ predictably according to the class of sensory stimulus provided (**Fig. 2a**). This result was not surprising since, in order to change its locomotion state, the animal is bound to produce a set of complementary movements and to adjust its posture accordingly (Machado et al., 2015).

The surprise came when we compared the level of stimulus-response specificity provided by a high dimensional description of mouse behaviour with locomotion alone. We found that the former increased the level of specificity and, in some conditions, locomotion largely failed to capture a substantial level of specificity that was instead clearly apparent in other behavioural

`

measures (**Fig. 3**). The increase in specificity was particularly remarkable when comparing behavioural responses to a loud sound and a visual looming, two instances of a defensive response in which, in absence of an escape shelter, the animal freezes (**Fig. 3a**). Linking the neural repertoire to the behavioural repertoire based on locomotion alone would indicate convergence – different sensory processes ultimately lead to one single action. Instead, by increasing the resolution of the behavioural repertoire, we were able to reject the convergence hypothesis showing that different sensory processes lead to different actions. In general a higher dimensionality always enabled a more accurate decoding of the stimuli based on the postures and the motor actions (**Fig. 3c,d**). At least part of this result can be explained by the fact that a higher dimensional behavioural quantification revealed a larger number of distinct behaviours that are specific for a particular class of stimuli (**Fig. 4**).

A clear outcome of our analyses, based on Mutual Information and clustering, was the fact that each stimulus class can evoke more than a single behavioural response (**Fig. 4c**). This is consistent with a previous result, obtained by using locomotion as unique behavioural descriptor, where looming was shown to evoke either simple immediate or escape followed by freeze (Shang et al., 2018). However our higher dimensional descriptors provide a substantially enhanced picture of this phenomenon indicating that the one-to-many map between stimuli and behavioural response occurs robustly across different classes of sensory stimuli. The benefit of expressing multiple responses to the same stimulus is not clear and most research so far has been focussing instead on the invariable traits of sensory guided behaviours. However this behavioural variability could be advantageous for survival since a unique response could make the animal too predictable and its behaviour easy to counteract. The neuronal substrates of this variability are still poorly understood and we believe will represent an important target for future studies.

`

Methods

Ethical Statement:

Experiments were conducted in accordance with the Animals, Scientific Procedures Act of 1986 (United Kingdom) and approved by the University of Manchester ethical review committee.

Animal Resources, Housing & Handling:

In this study we used C57Bl/6 mice (n = 29, all male) obtained from obtained from the Biological Services facility at University of Manchester, originally from Envigo. All mice were stored in cages of 3 individuals and were provided with food and water ad libitum. During transfer between the cage and the behavioural arena we used the tube handling procedure instead of tail picking, as prescribed in (Hurst and West, 2010), in order to minimise stress and reduce variability across animals. Mice were kept on a 12:12 light dark cycle. Behavioural testing was conducted at Zeitgeber time 6 or 18 (respectively n = 14 and 15 animals). All experiments were conducted under auditory white noise background at 64 dB(C).

Experimental Set-Up:

The animals were recorded in an open field arena with 4 programmable cameras (Chamaleon 3 from Point Grey; frame rate = 15Hz). The camera lenses were covered with infrared cut-on filters (Edmund Optics) and fed with constant infrared light. In order to deliver the flash stimulation we used two LEDs mounted inside the arena (model LZ4-00B208, LED engin; controlled by T-Cube drivers, Thorlabs). This stimulus provided diffuse excitation of all photoreceptors (S-cone opsin: $4.43*10^{12}$ photon/cm^2/s; Melanopsin: $2.49*10^{15}$ photon/cm^2/s; Rhodopsin: $1.98*10^{15}$ photon/cm^2/s; M-cone opsin: $7.09*10^{14}$ photon/cm^2/s). The looming stimuli were delivered by a projector onto a rear projection screen mounted at the top of the arena (background illumination: $4.08*10^{10}$, $1.65*10^{13}$, $1.94*10^{13}$ and $2.96*10^{13}$ photon/cm^2/s respectively S-cone opsin, Melanopsin, Rhodopsin and M-cone opsin). The auditory stimuli were provided by two speakers positioned outside the arena.
The experiments were controlled by using Psychopy (version 1.82.01) (Peirce, 2007). Frame acquisition was synchronized with the projected images and across cameras by a common electrical trigger delivered by an Arduino Uno board (arduino.cc) controlled by Psychopy through a serial interface (pyserial). Trigger control was enabled on Chamaleon 3 cameras through FlyCapture2 software (from Point Grey). All movies were encoded as M-JPEG from RGB 1280 (W) x 1040 (H) images. For tracking RGB images were converted to grayscale. Calculation of retinal irradiance for each photoreceptor was based on Govardovskii templates (Govardovskii et al., 2000) and lens correction functions (Jacobs and Williams, 2007). As auditory stimuli we used either a pure tone (C6 at 102 dB(C)) or a white noise (at 89 dB(C)) both presented for 1 second.

`

For all experiments we delivered 6 blocks of stimuli where each block was constituted by a flash, a looming and a sound. The order of the stimuli was independently randomised within each block. As looming stimulus we used two variants: a "standard" black looming (87% Michelson Contrast; looming speed = 66deg/s) and a modified looming where the black disc was replaced by a disc with a grating pattern (Spatial Frequency = 0.068 cycles/degree; Michelson Contrast: 35% for white vs grey, 87% for grey vs black, 94% for white vs black; looming speed = 66deg/s). As sound stimulus we used either a pure tone (1046.5 Hz, duration = 1s) or a white noise pitch (duration = 1s). The selection of looming and sound variants was randomly generated at each trial.

Reconstruction of three dimensional poses:

Three dimensional reconstruction of the mouse body was based on simultaneously tracked body landmarks from four the cameras (**Supplementary Fig. 1a,b**). The four camera system was calibrated using the Direct Linear Transform algorithm (Hartley, 2004) before data collection by using Lego® objects of known dimensions (**Supplementary Fig. 1c-f**). After data collection body landmarks were detected independently for each camera by using DeepLabCut software (Mathis et al., 2018). We used $n = 5$ body landmarks: the nose-tip, the left and right ears, the neck base and the tail base (as shown **Fig. 1a**). When the confidence of a landmark was higher than 0.5 the landmark was considered valid. Valid landmarks were then used to estimate the 3D coordinates of the body points using least square triangulation. The result of this initial 3D reconstruction was saved as raw reconstruction (**Fig. 1b**, left panel). The raw reconstruction contained outlier poses caused by incorrect or missing landmark detections (typically occurring when the relevant body parts were occluded). To correct those outliers we developed a method that automatically identifies correctly reconstructed body points and uses the knowledge of the geometrical relations between all points to re-estimate the incorrectly reconstructed (or missing) points. Knowledge of these geometrical relations was provided by a Statistical Shape Model (SSM).

We first estimated a statistical shape model (SSM) of the mouse body based on $n = 5$ body points (Cootes et al., 1995). This was achieved by using a set of 400 poses, each represented by a $n \times 3$ matrix $\boldsymbol{X}_{train}$ whose correct 3D reconstruction was manually assessed. Each training pose $\boldsymbol{X}_{train}$ was aligned to the reference pose $\overline{\boldsymbol{X}}$ using Partial Procrustes Superimposition (PPS). This algorithm estimates the 3×3 rotation matrix $\boldsymbol{R}$ and the $n \times 3$ translation $\boldsymbol{T}$ matrix that minimize the distance $\|\overline{\boldsymbol{X}} - (\boldsymbol{X}_{train}\boldsymbol{R} + \boldsymbol{T})\|_F$ calculated by using the Frobenius norm. A principal component analysis was then performed on the aligned poses to obtain a set of eigenposes $\mathbf{P}$ and eigenvalues λ. The first $p = 3$ eigenposes were sufficient to explain >80% of the shape changes in our training set. Based on those eigenposes the SSM model enabled to express any aligned pose $\boldsymbol{X}$ as

$$\boldsymbol{X} = \overline{\boldsymbol{X}} + \sum_{i}^{p} b_i \boldsymbol{P}_i \qquad (2)$$

`

where b_i represent the shape parameters. To identify outlier poses each pose $\boldsymbol{X}$ was first aligned to the reference pose $\overline{\boldsymbol{X}}$ and shape parameters were estimated. A pose was labelled as incorrect when either the Euclidean distance between $\overline{\boldsymbol{X}}$ and $\boldsymbol{X}$ or any of the shape parameters exceeded pre-set thresholds.

Outlier poses could be corrected if only 1-2 body points were incorrectly reconstructed by using the remaining body points and the trained SSM. Correctly reconstructed body points, represented by the $(n-2) \times 3$ matrix $\mathbf{X}_{subset}$, were identified as the subset of points, out of all possible $(n-2)$ subsets, that minimized the distance $\|\overline{\boldsymbol{X}}_{subset}\boldsymbol{R}_{subset} + \boldsymbol{T}_{subset} - \boldsymbol{X}_{subset}\|_F$. Here the matrices $\boldsymbol{R}_{subset}$ and $\boldsymbol{T}_{subset}$ were obtained by aligning the corresponding body points of the reference pose, $\overline{\boldsymbol{X}}_{subset}$, to the selected $(n-2) \times 3$ matrix $\boldsymbol{X}_{subset}$. The shape parameters b_i were treated and missing data and re-estimated by applying Piecewise Cubic Hermite Interpolation on the shape parameter time series. The corrected pose $\boldsymbol{X}$ was then re-estimated as $\boldsymbol{X} = \left(\overline{\boldsymbol{X}} + \sum_{i=1}^{p} b_i \boldsymbol{P}_i\right)\boldsymbol{R}_{subset} + \boldsymbol{T}_{subset}$.

These preliminary stages enabled to replace gross outliers in the raw 3D reconstruction. We then used all poses $\boldsymbol{X}$ and associated shape parameters as input for an optimization procedure aimed at obtaining a refined 3D reconstruction by minimizing the following cost function:

$$C(\boldsymbol{b}, \boldsymbol{R}, \boldsymbol{T}) = \left\| \overline{\boldsymbol{X}} - \left(\boldsymbol{X} + \sum_{i=1}^{p} b_i \boldsymbol{P}_i \right) \boldsymbol{R} + \boldsymbol{T} \right\|_F + \alpha \sum_{i=1}^{N_p} \frac{b_i^2}{\lambda_i} \qquad (3)$$

where the right-hand side of the *equation 3* represents a regularization factor to penalize for excessive changes in body shape. The value for the regularization parameter α, set at 0.001, was determined by first applying this cost function to a simulated dataset. For all further analyses the time series of each element of $\boldsymbol{b}$, $\boldsymbol{R}$ and $\boldsymbol{T}$ were smoothed using the kernel *w* = [0.2 0.6 0.2]. After smoothing each rotation matrix $\boldsymbol{R}(t)$ was renormalized by using Singular Value Decomposition.

Following this reconstruction procedure the mouse pose at any given frame t was defined by shape parameters $\boldsymbol{b}(t)$ and rigid transformations $\boldsymbol{R}(t)$ and $\boldsymbol{T}(t)$ as reported in *equation 1* and as shown in **Fig. 1b** (right panel).

Rank Estimation:

For rank estimation we used the Bi-Cross Validation method proposed by (Owen and Perry, 2009). The $m \times n$ response matrix $\boldsymbol{X}$ is partitioned into four submatrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ where $\boldsymbol{A} \in R^{r \times s}, \boldsymbol{B} \in R^{r \times (n-s)}, \boldsymbol{C} \in R^{(m-r) \times s}, \boldsymbol{A} \in R^{(m-r) \times (n-s)}$. Then the matrices $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ could be used to predict $\boldsymbol{A}$. Specifically if both $\boldsymbol{X}$ and $\boldsymbol{D}$ have rank k then $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{D}^{+}\boldsymbol{C} = \boldsymbol{B}\left(\widehat{\boldsymbol{D}}^{k}\right)^{+}\boldsymbol{C}$ (Owen and Perry, 2009), where $\boldsymbol{D}^{+}$ represents the pseudoinverse of $\boldsymbol{D}$ and $\widehat{\boldsymbol{D}}^{k}$ represents the *k*-rank approximation of D obtained by Singular Value Decomposition. Using this property we partitioned the rows and columns of $\boldsymbol{X}$ respectively into h and l subsets so

`

that each $h \times l$ subset represented a different hold out matrix $\boldsymbol{A}$. Finally we estimated the Bi-Cross Validation error as function of the *k*-rank approximation of the ***D*** matrices as:

$$BCV(k) = \sum_{i=1}^{h}\sum_{j=1}^{l}\left\|A_{i,j} - B_{i,j}\left(\widehat{D}_{i,j}^{(k)}\right)^{+} C_{i,j}\right\|_F^2 \qquad (4)$$

By systematically changing *k* we expect the error would reach its minimum around the true rank of $\boldsymbol{X}$.

Generation of Behavioural Measures and Dimensionality Reduction:

The full set of posture and movement measures was calculated from the refined 3D reconstruction as analytically described in **Fig. 1d**. Each measure was then quantile normalized in the range [0, 1].
For all estimates of Specificity Index, Stimulus Decoding and Behavioural Clustering both the full set of measures of body movements and postures and the locomotion only measures were pre-processed by Principal Component Analysis. The number of components was restricted to the minimum required to explain 80% variance of the original data.

Stimulus-Response Specificity and Decoding:

The Specificity Index (SI) for each behavioural response was estimated as the fraction nearest neighbour responses evoked by the same stimulus class. Performances for all the decoders were estimated by using 10-fold cross-validation. For robustness this estimate was repeated 50 times. K-Nearest Neighbour decoding was run by systematically varying the parameter K chosen in the range [1, 50]. Classification Tree decoding was based on a single tree with a maximum number of 20 branches. Random Forest decoding was performed by combining 100 classification trees, each limited to a maximum number of 20 branches.

Clustering and Information Analysis

Clustering was performed by using k-means algorithm with k-means++ initialization (Arthur and Vassilvitskii, 2007). The number of clusters k was systematically increased in the range (2-30). For each value of k, clustering was repeated 50 times and for each repeat the best clustering results was selected from 100 independent runs. In order to estimate Shannon's Mutual Information the probabilities distributions $p(G)$ and $p(G|S)$, where $\mathbb{G} = (g_1, \dots, g_k)$ indicates the cluster set and $\mathbb{S} = (s_1, \dots, s_n)$ the stimulus set, were estimated directly from the frequency histograms obtained from our dataset. From these distributions the response and noise entropies were calculated as

$$H(G) = -\sum_{g \in G} p(g) log_2 p(g)$$

`

$$H(G|S) = -\sum_{g \in G, s \in S} p(g,s) log_2 p(g|s)$$

These naïve estimates were then corrected for the sampling bias by using quadratic extrapolation as in (Strong et al., 1998). Mutual Information (MI) was then calculated from the difference of these corrected estimates. The change in MI as function of the number of clusters was fit with the function

$$MI(K) = a\left(1 - e^{\frac{-k}{\tau}}\right) + bk$$

through a mean square error minimization based on the interior point method (Matlab function fmincon). For this fitting the values of the parameters a, b and τ were constrained to be positive.

`

References

Arthur, D., and Vassilvitskii, S. (2007). k-means plus plus : The Advantages of Careful Seeding. Proceedings of the Eighteenth Annual Acm-Siam Symposium on Discrete Algorithms, 1027-1035.

Bernstein, J.G., and Boyden, E.S. (2011). Optogenetic tools for analyzing the neural circuits of behavior. Trends Cogn Sci 15, 592-600.

Cooke, S.F., Komorowski, R.W., Kaplan, E.S., Gavornik, J.P., and Bear, M.F. (2015). Visual recognition memory, manifested as long-term habituation, requires synaptic plasticity in V1. Nat Neurosci 18, 926-926.

Cootes, T.F., Taylor, C.J., Cooper, D.H., and Graham, J. (1995). Active Shape Models - Their Training and Application. Comput Vis Image Und 61, 38-59.

Davis, M., Falls, W.A., Campeau, S., and Kim, M. (1993). Fear-Potentiated Startle - a Neural and Pharmacological Analysis. Behav Brain Res 58, 175-198.

De Franceschi, G., Vivattanasarn, T., Saleem, A.B., and Solomon, S.G. (2016). Vision Guides Selection of Freeze or Flight Defense Strategies in Mice. Curr Biol 26, 2150-2154.

Evans, D.A., Stempel, A.V., Vale, R., Ruehle, S., Lefler, Y., and Branco, T. (2018). A synaptic threshold mechanism for computing escape decisions. Nature 558, 590-594.

Govardovskii, V.I., Fyhrquist, N., Reuter, T., Kuzmin, D.G., and Donner, K. (2000). In search of the visual pigment template. Visual Neurosci 17, 509-528.

Hartley, R.I.a.Z., A. (2004). Multiple View Geometry in Computer Vision, Second edn (Cambridge University Press).

Hoy, J.L., Bishop, H.I., and Niell, C.M. (2019). Defined Cell Types in Superior Colliculus Make Distinct Contributions to Prey Capture Behavior in the Mouse. Curr Biol 29, 4130-4138 e4135.

Hoy, J.L., Yavorska, I., Wehr, M., and Niell, C.M. (2016). Vision Drives Accurate Approach Behavior during Prey Capture in Laboratory Mice. Curr Biol 26, 3046-3052.

Hurst, J.L., and West, R.S. (2010). Taming anxiety in laboratory mice. Nat Methods 7, 825-U1516.

Jacobs, G.H., and Williams, G.A. (2007). Contributions of the mouse UV photopigment to the ERG and to vision. Doc Ophthalmol 115, 137-144.

Koch, M. (1999). The neurobiology of startle. Progress in neurobiology 59, 107-128.

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron 93, 480-490.

`

Machado, A.S., Darmohray, D.M., Fayad, J., Marques, H.G., and Carey, M.R. (2015). A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. eLife 4.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 21, 1281-1289.

Owen, A.B., and Perry, P.O. (2009). Bi-Cross-Validation of the Svd and the Nonnegative Matrix Factorization. Ann Appl Stat 3, 564-594.

Peirce, J.W. (2007). PsychoPy - Psychophysics software in Python. J Neurosci Meth 162, 8-13.

Reinhard, K., Li, C., Do, Q., Burke, E.G., Heynderickx, S., and Farrow, K. (2019). A projection specific logic to sampling visual inputs in mouse superior colliculus. eLife 8.

Shang, C.P., Chen, Z.J., Liu, A.X., Li, Y., Zhang, J.J., Qu, B.L., Yan, F., Zhang, Y.N., Liu, W.X., Liu, Z.H.*, et al.* (2018). Divergent midbrain circuits orchestrate escape and freezing responses to looming stimuli in mice. Nat Commun 9.

Shang, C.P., Liu, Z.H., Chen, Z.J., Shi, Y.C., Wang, Q., Liu, S., Li, D.P., and Cao, P. (2015). A parvalbumin-positive excitatory visual pathway to trigger fear responses in mice. Science 348, 1472-1477.

Smeds, L., Takeshita, D., Turunen, T., Tiihonen, J., Westo, J., Martyniuk, N., Seppanen, A., and Ala-Laurila, P. (2019). Paradoxical Rules of Spike Train Decoding Revealed at the Sensitivity Limit of Vision. Neuron 104, 576-587 e511.

Storchi, R., Rodgers, J., Gracey, M., Martial, F.P., Wynne, J., Ryan, S., Twining, C.J., Cootes, T.F., Killick, R., and Lucas, R.J. (2019). Measuring vision using innate behaviours in mice with intact and impaired retina function. Scientific reports 9, 10396.

Strong, S.P., Koberle, R., van Steveninck, R.R.D., and Bialek, W. (1998). Entropy and information in neural spike trains. Phys Rev Lett 80, 197-200.

Tovote, P., Esposito, M.S., Botta, P., Haudun, F.C., Fadok, J.P., Markovic, M., Wolff, S.B.E., Ramakrishnan, C., Fenno, L., Deisseroth, K.*, et al.* (2016). Midbrain circuits for defensive behaviour. Nature 534, 206-+.

Wei, P., Liu, N., Zhang, Z., Liu, X., Tang, Y., He, X., Wu, B., Zhou, Z., Liu, Y., Li, J.*, et al.* (2015). Processing of visually evoked innate fear by a non-canonical thalamic pathway. Nat Commun 6, 6756.

Yilmaz, M., and Meister, M. (2013). Rapid Innate Defensive Responses of Mice to Looming Visual Stimuli. Curr Biol 23, 2011-2015.