Cross-site Reproducibility of Social Deficits in Group-housed BTBR Mice Using Automated Longitudinal Behavioural Monitoring

Tatiana Peleh, a,1 Kevin G. O. Ike, b Ingeborg Frentz, b Bauke Buwalda, b Sietse F. de Boer, b Bastian Hengerer a and Martien J. H. Kas c

a* Boehringer Ingelheim Pharma GmbH & Co. KG, 88397 Biberach an der Riss, Germany
b Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

Abstract—Social withdrawal is associated with a variety of neuropsychiatric disorders, including neurodevelopmental disorders. Rodent studies provide the opportunity to study neurobiological mechanisms underlying social withdrawal, however, homologous paradigms to increase translatability of social behaviour between human and animal observation are needed. Standard behavioural rodent assays have limited ethological validity in terms of number of interaction partners, type of behaviour, duration of observation and environmental conditions. In addition, reproducibility of behavioural findings in rodents is further limited by manual and subjective behavioural scoring. Using a newly developed automated tracking tool for longitudinal monitoring of freely moving mice, we assessed social behaviours (approach, sniff, follow and leave) over seven consecutive days in colonies of BTBR and of C57BL/6J mice in two independent laboratories. Results from both laboratories confirmed previous findings of reduced social interaction in BTBR mice revealing a high level of reproducibility for this mouse phenotype using longitudinal colony assessments. In addition, we showed that detector settings contribute to laboratory specific findings as part of the behavioural data analysis procedure. Our cross-site study demonstrates reproducibility and robustness of reduced social interaction in BTBR mice using automated analysis in an ethologically relevant context.

This article is part of a Special Issue entitled: Animal Models. © 2020 The Author(s). Published by Elsevier Ltd on behalf of IBRO. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Key words: automated tracking, ethology, long-term monitoring, translational neuroscience, social behaviour, circadian.

INTRODUCTION

Social dysfunction, e.g. social withdrawal, is a common (early) symptom of many neuropsychiatric disorders including schizophrenia, major depression, Alzheimer’s disease and autism spectrum disorders (ASD) (American Psychiatric Association, 2013). Finding treatments for mental disorders is challenging due to the lack of biomarkers and poor understanding of the causal mechanisms. One confounding issue in mental health research is limited translatability between primarily subjective diagnostic criteria (DSM-V and ICD) and validation of animal models for uniquely human and highly diverse mental conditions. Novel research concepts like Research Domain Criteria (RDoC) focus on single domains and symptoms, e.g. social withdrawal, in order to understand the underlying biological mechanism of symptoms independent of diagnosis (Cuthbert and Insel, 2010, 2013). Following the concept of RDoC, the European PRISM (Psychiatric ratings using intermediate stratified markers) project aims to provide a transdiagnostic perspective on social dysfunction sharing common neurobiological mechanisms across schizophrenia, major depression and Alzheimer’s disease patients (Kas et al., 2019; Porcelli et al., 2019). In practice, social dysfunction is often measured by making use of questionnaires such as the social functioning scale (Birchwood et al., 1990). Recently, behavioural tasks (see for example Hynes et al., 2011) and portable measuring devices, such as smartphones, (see for example Chow et al., 2017; Eskes et al., 2016) have also been introduced to assess social behaviour. From these methods it has become clear that human social behaviour can be complex. Often it involves groups of people rather than interactions between just 2 peers (i.e. dyadic interactions). A key aspect of translatability is the alignment of the assessment of human (social) behaviour and rodent behavioural paradigms (Peleh et al., 2019b). However, current standard social behavioural tasks in mice, e.g. the three-chamber test, are simple and often only include dyadic interactions (Crawley, 2007; Moy et al., 2004; Nadler...
et al., 2004; Pearson et al., 2010). As such, the translational value of these tests is under discussion. More complex paradigms, such as the Visible Burrow System (VBS), enable longitudinal monitoring of groups of freely moving animals under naturalistic conditions (Arakawa et al., 2007; Blanchard et al., 1995, 2001; Pobbe et al., 2010). These ethologically relevant paradigms provide a comprehensive readout of the dynamics of social behaviour including social hierarchy formation and changes of social encounters (following, sniffing etc.) over time (day/night periods) (Arakawa et al., 2007; Blanchard et al., 1995, 2001; Pobbe et al., 2010). However, they primarily rely on the manual annotation of the displayed social behaviours, substantially increasing the time and work required for assessing social dysfunction compared to the current standard.

Recently, there has been an uprising in technological solutions aimed at automatically scoring and tracking the behaviour in groups of rodents (for review see Peleh et al., 2019b), making wide-scale use of complex behavioural paradigms such as the VBS a viable option. With automatic scoring it is possible to score social behaviour across the full 24 h of multiple days, giving a complete overview of the circadian activity of the animal. In contrast, previous studies employing similar paradigms were forced to take snapshots of the behaviour, only analysing specific time-windows the researchers deemed crucial (see for example Bove et al., 2018). A recent study has shown that time of testing (dark or light period) has a significant impact on social behaviour in C57BL/6N mice (Richetto et al., 2019). Moreover, studies involving genetic deletions related to neuropsychiatric disorders have shown that these genes can influence the circadian activity pattern (Maple et al., 2018). Together this suggests that key differences between animals can be misinterpreted when only looking at snapshots of their behaviour across the day. Not only time can be a crucial factor leading to inconsistencies in behavioural findings. Next to time, housing conditions, experimenter bias and duration of testing and subjective behavioural scoring have all been found to lead to inconsistencies (Hurst and West, 2010; Richter et al., 2010; Wahlsten et al., 2003). Many of these factors can be tackled by making use of longitudinal automatic behavioural scoring (for review see Peleh et al., 2019b).

In the study presented here, we aim to assess the reproducibility of long-term automatic behavioural scoring of social behaviours across the full circadian period in a complex, ethologically relevant, behavioural paradigm. For this, we assessed and compared social behaviours in the BTBR T + Itpr3tf/J (BTBR) and C57BL/6J inbred mouse strains in two independent laboratories using a recently developed automated tracking tool for group-housed mice (Peleh et al., 2019a). The BTBR mouse is known for its reproducible social deficits and repetitive behaviours demonstrated using standard behavioural assays (McFarlane et al., 2008; Moy et al., 2007; Yang et al., 2012). Long-term observations of BTBR mice in the VBS confirmed consistency of social deficits during particular day and light periods (Bove et al., 2018; Pobbe et al., 2010). Thus, we expect to find comparable social deficiencies when assessing its behaviour across the full 24 h of each day.

**EXPERIMENTAL PROCEDURES**

**Animals**

Inbred male C57BL/6J (JAX stock #000664) and BTBR T + Itpr3tf/J (JAX stock #002282) mice were ordered from The Jackson Laboratory (Bar Harbor, Maine, USA) via Charles River Europe (Den Bosch, The Netherlands), by both experimental sites (i.e. University of Groningen and Boehringer Ingelheim Pharma GmbH & Co. KG). After arrival, a minimum of two weeks before the start of the experiments, mice were housed in groups of 2–4 in Makrolon type 3 cages. Animals were subjected to site-specific standard housing conditions. Mice were maintained under a 12:12 light/dark cycle, controlled temperature (21 ± 1 °C) and humidity (55 ± 5%) and with ad libitum access to food and water. All animals were kept on site-specific standard bedding (Aspen or Corncob based), and had access to shredded cardboard nesting material, a cardboard tube and a red plastic igloo as enrichment. At the start of the experiment, all mice were 12–20 weeks of age (adult). Experiments were conducted following the approval of the responsible animal ethics committees. All animal care

**Table 1. Differences in detector settings used for automated behavioural recognition between the two experimental sites**

<table>
<thead>
<tr>
<th></th>
<th>Lab A</th>
<th>Lab B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social sniffing:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance between mice</td>
<td>&lt;1 cm</td>
<td>&lt;3.5 cm</td>
</tr>
<tr>
<td>Minimum duration</td>
<td>0.5 s</td>
<td>0.33 s</td>
</tr>
<tr>
<td><strong>Social approach:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance between mouse</td>
<td>&lt;20 cm</td>
<td>&lt;100 cm</td>
</tr>
<tr>
<td>Moving direction (angle) of mouse 1</td>
<td>&lt;30°</td>
<td>&lt;45°</td>
</tr>
<tr>
<td>Distance to be travelled by mouse 1 towards mouse 2</td>
<td>&gt;10 cm</td>
<td>&gt;7 cm</td>
</tr>
<tr>
<td>Velocity of mouse 1</td>
<td>&gt;3 cm/ s</td>
<td>&gt;5 cm/s</td>
</tr>
<tr>
<td><strong>Social leave:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance between mouse</td>
<td>&lt;10 cm</td>
<td>&lt;100 cm</td>
</tr>
<tr>
<td>Moving direction (angle) of mouse 1</td>
<td>&gt;30°</td>
<td>&gt;45°</td>
</tr>
<tr>
<td>Distance to be travelled by mouse 1 away from mouse 2</td>
<td>&gt;10 cm</td>
<td>&gt;14 cm</td>
</tr>
<tr>
<td>Velocity of mouse 1</td>
<td>&gt;3 cm/ s</td>
<td>&gt;1 cm/s</td>
</tr>
<tr>
<td><strong>Social follow:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance between mouse</td>
<td>&lt;20 cm</td>
<td>&lt;30 cm</td>
</tr>
<tr>
<td>Moving direction (angle) of mouse 1</td>
<td>&lt;30°</td>
<td>&lt;45°</td>
</tr>
<tr>
<td>Moving direction (angle) of mouse 2</td>
<td>&gt;90°</td>
<td>&gt;90°</td>
</tr>
<tr>
<td>Distance to be travelled by mouse 1 and mouse 2</td>
<td>&gt;10 cm</td>
<td>&gt;7 cm</td>
</tr>
<tr>
<td>Velocity of mice</td>
<td>&gt;4 cm/ s</td>
<td>&gt;1 cm/s</td>
</tr>
</tbody>
</table>
136 was performed according to the local rules set by site-
137 specific ethical authorities.

**Experimental conditions**

All procedures, including animal handling and animal
holding were aligned across laboratories to the extent of
feasibility. At least two weeks before the start of the
experiment all animals received an “ear-cut” to allow for
individual identification of the group housed animals.
During the weeks preceding the experiment all cages
were cleaned and all animals were weighed and
handled weekly. One week before starting the
experiment, animals were injected with a Radio-
frequency identification (RFID) chip (8 mm × 2 mm).
Injection of the chip took place under anaesthesia (i.e.
Isoflurane ~ 2%). For chip injection, the dorsal/caudal
part of the animal was cleaned with a disinfecting agent
(e.g. 70% ethanol) and, subsequently, the RFID-chip
was injected subcutaneously with a specialised syringe
just right of the median on the dorsal/caudal part of
the mouse. This leads to a secure upright position of
the chip under the skin. At the first experimental day, 4
animals were placed at random in a semi-natural
environment starting at 1 h before lights-off. The animals
placed in one environment were taken from different
cages, thus the animals placed together did not interact
in the two weeks before placement. During the
recordings for seven consecutive days animals were left
undisturbed with the exception of short daily checks for
animal welfare. All animal handling was conducted while
wearing gloves.

**Fig. 1.** Reduced social approach behaviour in BTBR mice measured in laboratories A and B with site-specific detector settings (DS). (A) Time spent
approaching per 12 h during dark (left) and light phase (right) in BTBR and C57BL6/J mice in laboratory A. Mixed effects model for RM followed by
Bonferroni, \( p < 0.0001 \) vs. C57BL/6J for dark phase and \( p < 0.0001 \) for light phase. (B) Time spent approaching per 12 h during dark (left) and light
phase (right) in BTBR and C57BL6/J mice in laboratory B. Mixed effects model for RM followed by Bonferroni, \( p < 0.0001 \) vs. C57BL/6J for dark
phase and \( p < 0.0001 \) for light phase. Data are expressed as boxplots with maximum and minimum values and a line at the median (Laboratory A:
\( n = 16 \) for BTBR group, \( n = 16 \) for C57BL/6J group, Laboratory B: \( n = 16 \) for BTBR group, \( n = 12 \) for C57BL/6J group).
The behavioural setup consisted of one open arena (80×60×50 cm) connected to four small nestboxes (7×7×7 cm) and one big nestbox (10×10×10 cm) via short tunnels (4×7 cm) through which the animals can freely pass. Two food hoppers and two water bottles were connected to the open arena to which animals had ad libitum access. The open arena was filled with approximately 1–2 cm of standard local bedding material. Each open arena was equipped with a translucent Perspex cover. On the top of this cover two sets of LEDs have been attached to provide equal illumination of the open arena. One set provides white light at a 12:12 light/dark-cycle, matching that of regular housing. The other provides infrared (IR) illumination to allow for 24 h recording. Recording is done using a Basler acA1300-60gmNIR GigE monochrome camera equipped with an 850 nm IR pass filter, mounted above the setup. The open arena rests on a plate containing 24 RFID antennae to allow for registration of the RFID chips. Behavioural characterisation of a combination of video and RFID signals was done utilizing RFID-assisted Socialscan (Clever Sys Inc., Reston, Virginia, USA) (Peleh et al., 2019a). In the first part of data analysis, behavioural scoring was performed independently applying site-specific detector settings (DS) (in the following referred to as “DS A” for laboratory A and “DS B” for laboratory B) based on manual annotation. The correlation between manual and automatic behavioural annotation, for both laboratory A and B, can be found in supplementary Fig. 1. DS included parameters such as moving direction (angle), distance between animals, and...
Distance moved for automated behavioural analysis of social approaching, social sniffing, social following and social leaving (Table 1). In the second part of data analysis, DS were aligned across laboratories by applying DS B on data from laboratory A. Then, combined data from both laboratories with aligned DS was evaluated in terms of influencing factors including strain, time and laboratory site.

**Statistical analysis**

Duration of each behaviour analysed per dark (12 h) and per light phase (12 h) for each day was analysed using a mixed effects model for repeated measures (RM) followed by Bonferroni’s post-hoc test. Results are expressed as mean ± S.E.M. Statistical analyses were performed using Graph Pad 8.0 (GraphPad Software, San Diego, CA) for Windows. Differences were considered statistically significant when p value was less than 0.05. The analysis of the combined data was performed on 12 h bins by means of a linear mixed effects regression model in R (R Core Team, 2018), making use of the packages “lme4” (Bates et al., 2015), “lmerTest” (Kuznetsova et al., 2017) and “emmeans” (Lenth, 2019). A log transformation was applied to approach a normal distribution.

**RESULTS**

**Reproducibility of social deficits in group-housed BTBR mice measured automatically at two laboratories with site-specific detector settings**

Consistent with previous observations, group-housed BTBR mice displayed reduced social behaviours.
Time spent leaving in laboratory A with DS A

(A) Time spent leaving per 12 h during dark (left) and light phase (right) in BTBR and C57BL/6J mice in laboratory A. Mixed effects model for RM followed by Bonferroni, $p = 0.0010$ vs. C57BL/6J for dark phase and $p < 0.0001$ for light phase. (B) Time spent leaving per 12 h during dark (left) and light phase (right) in BTBR and C57BL/6J mice in laboratory B. Mixed effects model for RM followed by Bonferroni, $p < 0.0001$ vs. C57BL/6J for dark phase and $p < 0.0001$ for light phase. Data are expressed as boxplots with maximum and minimum values and a line at the median (Laboratory A: $n = 16$ for BTBR group, $n = 16$ for C57BL/6J group, Laboratory B: $n = 16$ for BTBR group, $n = 12$ for C57BL/6J group).

Time spent leaving in laboratory B with DS B

Fig. 4. Reduced social leaving behaviour in BTBR mice measured in laboratories A and B with site-specific detector settings (DS). (A) Time spent leaving per 12 h during dark (left) and light phase (right) in BTBR and C57BL/6J mice in laboratory A. Mixed effects model for RM followed by Bonferroni, $p = 0.0010$ vs. C57BL/6J for dark phase and $p < 0.0001$ for light phase. (B) Time spent leaving per 12 h during dark (left) and light phase (right) in BTBR and C57BL/6J mice in laboratory B. Mixed effects model for RM followed by Bonferroni, $p < 0.0001$ vs. C57BL/6J for dark phase and $p < 0.0001$ for light phase. Data are expressed as boxplots with maximum and minimum values and a line at the median (Laboratory A: $n = 16$ for BTBR group, $n = 16$ for C57BL/6J group, Laboratory B: $n = 16$ for BTBR group, $n = 12$ for C57BL/6J group).
259 = 87.8, \( p < 0.0001 \) by mixed effects model for RM followed by Bonferroni multiple comparison test, Fig. 3 B). By the end of the monitoring period, difference in following duration between BTBR and C57BL6/J mice vanished in both laboratories (Fig. 3 A, B). Time spent leaving was significantly reduced during dark and light phases both in laboratory A (dark: \( F(1, 30) = 13.2, p = 0.0010 \); light: \( F(1, 30) = 48.2, p < 0.0001 \) by mixed effects model RM followed by Bonferroni multiple comparison test, Fig. 4 A) and in laboratory B (dark: \( F(1, 26) = 23.2, p < 0.0001 \); light: \( F(1, 26) = 169, p < 0.0001 \) by mixed effects model RM, Fig. 4 B). In laboratory A, however, time spent leaving was significantly different between BTBR and C57BL6/J mice during dark phase 1 and 2 (\( p < 0.01 \)) but not during the remaining dark phases (Fig. 4 A). In laboratory B, leaving duration was significant during dark phase 2 to 6 but not during dark phase 1 and 7 (Fig. 4 B).

Laboratory differences in detector settings

Interestingly, absolute duration of social behaviours measured in laboratory B with site-specific DS was generally higher when compared to laboratory A. In order to assess the effect of detector settings (DS A vs DS B) on behavioural data, we compared time spent approaching in C57BL6/J mice in laboratory A with site-specific detector settings DS A (in blue) and with DS B (in orange). Two-way Anova for RM followed by Bonferroni, \( p < 0.0001 \) vs. DS B for dark phase and \( p = 0.0014 \) for light phase. Data are expressed as boxplots with maximum and minimum values and a line at the median and contain data set from laboratory A analysed with DS A (n = 16 mice) and DS B (n = 16 mice).

Fig. 5. Site-specific detector settings produced different absolute values of time spent approaching in C57BL6/J mice. (A) Time spent approaching in C57BL6/J mice per hour over 168 consecutive hours measured in laboratory A with site-specific detector settings DS A (in red) and with DS B (in black). (B) Time spent approaching per 12 h during dark (left) and light phases (right) in C57BL6/J mice in laboratory A with site-specific detector settings DS A (in blue) and with DS B (in orange). Two-way Anova for RM followed by Bonferroni, \( p < 0.0001 \) vs. DS B for dark phase and \( p = 0.0014 \) for light phase. Data are expressed as boxplots with maximum and minimum values and a line at the median and contain data set from laboratory A analysed with DS A (n = 16 mice) and DS B (n = 16 mice).
Marginal interaction effects of strain, time and laboratory

In order to evaluate the effect of strain, time and laboratory on behavioural readouts, we aligned the
detector settings across laboratories (detector settings from laboratory B were applied on both data sets) and
in order to evaluate the effect of strain, time and
not light phases (dark: F (6, 180) = 5.5, p < 0.0001 and
light: F (6, 180) = 0.40, p = 0.8798 by two-way Anova
for RM). In particular, DS A produced significantly lower
absolute values (duration of social approaching) in
comparison to DS B on day 1 and day 2 during dark
(p = 0.0002 and p = 0.0205, respectively) and light
periods (p = 0.0255 and p = 0.0077, respectively, by
post-hoc Bonferroni’s multiple comparisons test, Fig. 5B).

DISCUSSION

The present study was designed to demonstrate cross-
site reproducibility of long-term automated scoring and
comparing of social behaviours in group-housed BTBR
containing data sets from laboratory A with DS B and
mice maintained in ethologically relevant
comparing of social behaviours in group-housed BTBR
and C57BL/6J mice maintained in ethologically relevant
conditions. Both laboratories reproduced social deficits
in BTBR mice consistent with previous findings of
reduced social approaching, sniffing and following
behaviour using manually scored dyadic interactions
(Bolivar et al., 2007; McFarlane et al., 2008; Moy et al.,
2019). The relationship between the response variables
in bins of 12 hour (i.e. time spent approach, sniffing, fol-
lowing and leaving) and strain of the animal, time and lab-
atory location (fixed effects), and all interactions
between these factors, was tested, while the individual
animal was added as random effect. Response variables
were transformed using a log transformation (i.e. log(x
+ 1)). For all behaviours, main effects of strain and time
were found (p < 0.0001). Main effect of location was sig-
nificant for approach and sniffing behaviour (p < 0.0001)
(Figs. 7 and 10). The two-way interaction between strain
and location was significant for time spent following
(p = 0.04) (Fig. 8). For time spent sniffing and following,
the two-way interaction between time and location was
significant (p < 0.001 and p = 0.01, respectively). Fur-
thermore, the three-way interaction between strain, time
and location was significant for time spent leaving
(p < 0.05) (Fig. 9). Post-hoc analysis, using Tukey cor-
rection for multiple comparison, was performed to assess
differences between the strains at specific time-points and
to determine how these differences varied between
laboratories.

In a day-by-day comparison, we observed temporal
changes in social behavioural elements. For example,
on the first day of the recording, social interactions were
increased when compared to the subsequent days and
are likely reflecting novelty-induced behavioural
responses. In addition, following day 4, an additional
temporal change in social behaviour was observed with
BTBR mice increasing and C57BL6/J mice maintaining
or slightly decreasing their social activities resulting
comparable levels of social activities in the two strains

Marginal interaction effects of strain, time and
laboratory

In order to evaluate the effect of strain, time and
laboratory on behavioural readouts, we aligned the
detector settings across laboratories (detector settings from laboratory B were applied on both data sets) and
and laboratory B with aligned detector settings. Time spent approach-
ing (A), sniffing (B), following (C) and leaving (D) per hour in BTBR
mice over 168 consecutive hours. Data are expressed as mean
± SEM and contain data sets from laboratory A with DS B and
laboratory B with DS B with a total n = 32 for BTBR group, n = 28 for
C57BL/6J group.

Fig. 6. Social behaviours in combined data sets from laboratory A and laboratory B with aligned detector settings. Time spent approach-
ing (A), sniffing (B), following (C) and leaving (D) per hour in BTBR
mice over 168 consecutive hours. Data are expressed as mean
± SEM and contain data sets from laboratory A with DS B and
laboratory B with DS B with a total n = 32 for BTBR group, n = 28 for
C57BL/6J group.
during the dark phase on day 7. Given the fact that both laboratories observed the same phenomenon, we exclude effects of laboratory environment as potential influencing factors for this temporal change in behaviour. Instead, it is tempting to speculate that these temporal changes in social behaviour are indicators for social hierarchy formation, which is established within the first 2–4 days (Williamson et al., 2016). However, further studies are required to assess the social hierarchy formation in these mice.

Differences between laboratories were observed in total duration of social behaviours. For example, while both laboratories clearly observed the social deficits in BTBR when compared to the C57BL/6J mice, the absolute behavioural scores in laboratory B were generally higher compared to laboratory A. Here we showed that these inter-laboratory differences can be explained, in part, by the use of site-specific detector settings determining duration of behavioural events.

Social approach, for instance, is defined, among others, by the moving direction expressed as the angle between the centre of the mouse being approached and the moving direction of the approaching mouse (for details see Peleh et al., 2019a). Setting this angle to 0 degree presupposes that the approaching mouse moves straight towards its conspecific while an angle of 45 degree allows the approaching mouse to move approximately into the direction of its conspecific. A wider angle here made the...
391 cut-off for approach behaviour less strict, leading to more
392 approach events and a longer duration of these events.
393 This same principle holds for many of the "detector set-
394 tings" used for the recognition of our chosen behaviours.
395 As the "strictness" of the detector settings of the two lab-
396 oratory sites may differ, this could partly explain why dif-
397 ferences in behaviour are found while making use of the
398 site-specific detector settings.
399 After aligning of the settings, the laboratory location
400 still showed a slight significant effect on the behavioural
401 findings, specifically in the latter part of the seven days
402 recordings. Closer inspection of the individual colony
403 data suggests that differences between the colonies
404 appear to be greater in one lab when compared to the
405 other. This was particularly the case in the second part
406 of the seven days recordings and coincided with the
407 time point during which the discrepancies between the
408 two locations were observed. Additional studies are
409 needed to investigate whether differences between
410 colonies are due to hierarchical structure variability
411 between the colonies (e.g., due to colony composition
412 based on prior animal history), to environmental factors
413 other than the ones controlled for in this study, or the
414 combination of these two. Nevertheless, the present

**Fig. 8.** Time spent following in bins of 12 h over 7 consecutive days, with aligned detector settings (DS B). (A) Combined data set, expressed as time spent approaching per 12 h in boxplots, with maximum and minimum values and a line at the median, n = 32 for BTBR and n = 28 for C57BL/6J. (B) data from laboratory A and laboratory B separated, expressed as time spent approaching per 12 h in boxplots, with maximum and minimum values and a line at the median, n = 16 for BTBR and n = 16 for C57BL/6J for lab A, and n = 16 for BTBR and n = 12 for C57BL/6J for lab B. Dark phase indicated by grey area and light phase by yellow area. Statistical analysis was performed on log-transformed data (i.e., log(x + 1)) by means of *linear mixed effects regression*, followed by post-hoc analysis with Tukey correction. *p < 0.05. See supplementary tables 3 and 4 for statistical values for post-hoc analysis.
415 study showed that longitudinal automated behavioural
416 observations in group housed mice can reliably detect
417 social deficits in BTBR mice in two different laboratory
418 environments, demonstrating the robustness of this
419 method.
420 The study’s colony-based approach provides many
421 advantages when compared to more conventional
422 methods aimed at detecting social behaviour. The effect
423 of the experimenter (see for example Bohlen et al.,
424 2014 or Sorge et al., 2014 ), both during the experiment
425 and in the subsequent analysis, is greatly reduced. There
426 is no human/animal interaction for the duration of the
427 experiment and data is analysed by means of an objective
428 algorithm, insensitive to subjective judgement or fatigue.
Concerning the behaviour itself, although similar social
deficits can be found, the current approach allows for
exploration of a multitude of behavioural phenotypes. As
such, social deficits in strains not displaying a robust
behavioural phenotype in conventional tasks, such as
the fmr1 KO mouse model (e.g. Sinclair et al., 2017;
Yau et al., 2016 ), focusing on a single aspect of social
behaviour might be elucidated in the current setup. The
fact that the present method also allows for longitudinal
testing of the social phenotype adds the possibility to dis-

Fig. 9. Time spent leaving in bins of 12 h over 7 consecutive days, with aligned detector settings (DS B). (A) Combined data set, expressed as time spent approaching per 12 h in boxplots, with maximum and minimum values and a line at the median, $n = 32$ for BTBR and $n = 28$ for C57BL/6J. (B) data from laboratory A and laboratory B separated, expressed as time spent approaching per 12 h in boxplots, with maximum and minimum values and a line at the median, $n = 16$ for BTBR and $n = 16$ for C57BL/6J for lab A, and $n = 16$ for BTBR and $n = 12$ for C57BL/6J for lab B. Dark phase indicated by grey area and light phase by yellow area. Statistical analysis was performed on log-transformed data (i.e. $\log(x + 1)$) by means of linear mixed effects regression, followed by post-hoc analysis with Tukey correction. *$p < 0.05$. See supplementary tables 5 and 6 for statistical values for post-hoc analysis.
tistinguish novelty induced social deficits from (innate) base-
line social differences. However, to fully demonstrate the
effectiveness of this relatively new method of behavioural
assessment further experiments have to be conducted,
making use of a variety of methods of manipulation (e.g.
by pharmacological or genetic modification). Given the
relative novelty of the current group-housed longitudinal
monitoring approach, sample size estimation is particu-
duarily difficult as there is hardly any data available yet to
base a reliable power-analysis estimation on. As such,
the most comparable literature is employing conventional
dyadic behavioural tasks, assessing the behavioural phe-
notype for minutes instead of days. In these experiments
social behaviour is often measured as time spent interact-
ing with a certain other individual (e.g. Moy et al., 2004).
However, in a colony setting this becomes a more compli-
cated setup, certainly in situations where genetic models
may be studied in groups varying in genotype (e.g. a col-
ony consisting of 1 knock-out animal and 3 wildtypes). In
a colony situation not all animals, in the same colony, are
exposed to the same situation. Controlling for this may
create the need to inflate sample size substantially, when
basing the expected effect on that found in conventional
setups. However, as a wider range of behaviours can

Fig. 10. Time spent sniffing in bins of 12 h over 7 consecutive days, with aligned detector settings (DS B). (A) Combined data set, expressed as
time spent approaching per 12 h in boxplots, with maximum and minimum values and a line at the median, n = 32 for BTBR and n = 28 for C57BL/6J.
(B) data from laboratory A and laboratory B separated, expressed as time spent approaching per 12 h in boxplots, with maximum and minimum
values and a line at the median, n = 16 for BTBR and n = 16 for C57BL/6J for lab A, and n = 16 for BTBR and n = 12 for C57BL/6J for lab B. Dark
phase indicated by grey area and light phase by yellow area. Statistical analysis was performed on log-transformed data (i.e. log(x + 1)) by means
of linear mixed effects regression, followed by post-hoc analysis with Tukey correction. *p < 0.05. See supplementary tables 7 and 8 for statistical
values for post-hoc analysis.
be observed, over a longer period of time, it is also more likely to pick-up a behavioural phenotype which would be missed in conventional behavioural tasks. Together this illustrates the need for future research to dive deeper into the effects of the group on the behaviour of the individual.

ACKNOWLEDGEMENTS

The PRISM project (www.prism-project.eu) leading to this application has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115916. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA. This publication reflects only the authors’ views neither IMI JU nor EFPIA nor the European Commission are liable for any use that may be made of the information contained therein. The authors would also like to thank Niels Jongens for his input on the statistical analysis.

REFERENCES


APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroscience.2020.04.045.