



OPEN

## Automation at the service of the study of executive functions in preclinical models

Francesca Zoratto<sup>1,6</sup>, Edoardo Pisa<sup>1,6</sup>, Claudia Soldati<sup>1</sup>, Caterina Barezzi<sup>1</sup>, Angela Maria Ottomana<sup>1,2</sup>, Martina Presta<sup>1,3</sup>, Valerio Santangelo<sup>4,5</sup> & Simone Macri<sup>1</sup>✉

Cognitive flexibility involves the capability to switch between different perspectives and implement novel strategies upon changed circumstances. The Wisconsin Card Sorting Test (in humans) and the Attentional Set-Shifting Task (ASST, in rodents) evaluate individual capability to acquire a reward-associated rule and subsequently disregard it in favour of a new one. Both tasks entail consecutive stages wherein subjects discriminate between: two stimuli of a given category (simple discrimination, SD); the stimuli of SD confounded by an irrelevant stimulus of a different category (compound discrimination, CD); different stimuli belonging to the SD category (intradimensional shift, IDS); and two stimuli of the confounding category (extradimensional shift, EDS). The ASST is labour intensive, not sufficiently standardised, and prone to experimental error. Here, we tested the validity of a new, commercially available, automated version of ASST (OPERON) in two independent experiments conducted in: different mouse strains (C57BL/6 and CD1 mice) to confirm their differential cognitive capabilities (Experiment 1); and an experimental model of chronic stress (administration of corticosterone in the drinking water; Experiment 2). In both experiments, OPERON confirmed the findings obtained through the manual version. Just as in Experiment 1 both versions captured the deficit of C57BL/6 mice on the reversal of the CD (CDR), so also in Experiment 2 they provided analogous evidence that corticosterone treated mice have a remarkable impairment in the IDS. Thus, OPERON capitalises upon automated phenotyping to overcome the limitation of the manual version of the ASST while providing comparable results.

Executive functions (EFs) encompass a variety of heterogeneous cognitive processes involved in the control and monitoring of other—lower-level—cognitive functions<sup>1</sup>. The core components of EFs<sup>2</sup> include the capability to inhibit behaviours and actions that are no longer appropriate to the purpose<sup>3</sup>; working memory, which is essential for the online maintenance of current goals<sup>4,5</sup>; and cognitive flexibility, i.e., the ability to vary and update effectively one's goals and objectives<sup>6</sup>. EFs allow mammals to pay attention to relevant cues and disregard meaningless information, exert inhibitory control over actions, and take optimal decisions<sup>7</sup>. It is important to emphasize that these processes appear to be consistently supported by the frontal lobes, another important aspect to consider for enclosing these heterogeneous capabilities into a single set of control functions<sup>8</sup>. In particular, cognitive flexibility, which involves the capacity to rapidly switch between different perspectives and implement novel strategies to cope with adverse situations or changed circumstances<sup>9</sup>, is thought to be supported by the dorsolateral portions of the prefrontal cortex<sup>10</sup>. Lesions to this area produce severe deficits ascribed to the so-called dysexecutive syndrome<sup>11</sup>. A main feature of dysexecutive syndrome is the absence of cognitive flexibility, evidenced by the perseverative behaviour with which these patients continue to choose a previously learned rule, as revealed by the Wisconsin Card Sorting Test (WCST, e.g.<sup>12</sup>). The WCST, commonly used to investigate cognitive flexibility in humans<sup>13</sup>, consists of a set of cards—varying for geometric shapes, colours and numbers—that must be sorted by the participant according to categories that change periodically. In order to identify the correct category, participants need to adjust card sorting on the basis of the examiner's positive or negative feedback (i.e., the chosen category is correct or incorrect, respectively), which follows any card sort.

<sup>1</sup>Centre for Behavioural Sciences and Mental Health, Istituto Superiore di Sanità, Rome, Italy. <sup>2</sup>Neuroscience Unit, Department of Medicine, University of Parma, Parma, Italy. <sup>3</sup>Department of Physiology and Pharmacology "Vittorio Ersamer", "Sapienza" University of Rome, Rome, Italy. <sup>4</sup>Department of Philosophy, Social Sciences and Education, University of Perugia, Perugia, Italy. <sup>5</sup>Functional Neuroimaging Laboratory, IRCCS Santa Lucia Foundation, Rome, Italy. <sup>6</sup>These authors contributed equally: Francesca Zoratto and Edoardo Pisa. ✉email: simone.macri@iss.it

The WCST specifically highlights the inability of patients with frontal lobe impairments to inhibit the previously learned rule and to update it in favour of a new rule that would allow to achieve the intended purpose<sup>14</sup>. Just as the WCST, the Cambridge Neuropsychological Test Automated Battery Intra-/Extra-Dimensional set-shift task (CANTAB ID/ED task<sup>15,16</sup>) evaluates rule acquisition and rule update focusing on EFs and can be considered a computerized analogue of the WCST.

Mental disorders, be them due to brain injuries or neuropsychiatric pathologies, are often associated with impaired EFs (DSM-5<sup>17</sup>). While representing a diagnostic feature in major and mild neurocognitive disorder (e.g., due to Alzheimer's disease, traumatic brain injury, HIV infection, Huntington's disease and substance-use disorder), compromised EFs are also an associated symptom in attention-deficit/hyperactivity disorder (ADHD) and schizophrenia (DSM-5). These disorders entail behavioural, neuroanatomical and physiological symptoms, and molecular, genetic and environmental causative factors. To detail the fundamental mechanisms underlying multifactorial neuropsychiatric disorders, experimental animal models have traditionally constituted a tool of choice in biomedical sciences. Within this framework, several authors designed and developed experimental models characterised by alterations at the level of the brain structures involved in EFs<sup>18,19</sup>. These have taken the form of lesion studies in which selected brain regions have been experimentally ablated (e.g.<sup>20,21</sup>), genetic engineering approaches targeting relevant molecular pathways (e.g.<sup>22,23</sup>), and environmental manipulations aimed at altering specific neurotransmitters (e.g.<sup>24,25</sup>), to name a few. Whilst these approaches reliably target physiological and neurological mediators of EFs, the paradigms to investigate this higher-order brain function in rodents are still scarce. Therefore, devising off-the-shelf tools to quantify EFs in laboratory animals represents a desirable goal holding promise to expand the translational value of preclinical studies of mental disorders. With respect to EFs, the Attentional Set-Shifting Task (ASST), originally developed by Birrell and Brown<sup>20</sup>, represents a welcome endeavour, whereby it translates the WCST to rodents, thus allowing the study of inhibitory control and cognitive flexibility<sup>26</sup>.

Similar to the WCST, the ASST evaluates individual capability to acquire a rule in order to obtain a reward, and to subsequently disregard such rule in favour of a new one, once the learning criterion has been attained by trial and error. The similarities between the WCST and the ASST are numerous: the reward of the WCST (experimenter feedback) is replaced in rodents by a palatable food reward located within a bowl underneath a digging medium; the dimensions of the deck of cards (geometric shapes, colours and numbers) are translated into odour, digging medium, and texture of the bowls, three highly ethologically-relevant dimensions in rodents; both tasks entail several stages assessing behavioural flexibility and inhibitory control with variable levels of difficulty. For example, with respect to the latter, in the transition between the simple discrimination and the compound discrimination (see the “Methods” section) the experimental subjects have to acquire a rule and follow such rule while disregarding a new dimension which is included as a confound. Afterwards, when requested to perform the reversal discrimination (see the “Methods” section) they have to inhibit their acquired response in favour of a new one which was previously incorrect. Finally, in the extra-dimensional shift stage, mice display their cognitive flexibility.

Beside its face validity, this task has proven elevated construct (sensitivity to procedures aimed at altering the biological substrates involved) and predictive (sensitivity to pharmacological treatments) validities. With respect to the former, Birrell and Brown<sup>20</sup> demonstrated that lesions of the frontal lobes in rats relate to deficits in the ASST; likewise, Hauser and co-authors<sup>27</sup> demonstrated that exposure to neonatal milk devoid of specific nutrients involved in neuronal patterning results in impaired EFs in the ASST. With respect to the latter, while we recently observed that methylphenidate administration (a compound used to mitigate attentional deficits in ADHD patients) improved mouse performance in the ASST<sup>28</sup>, Bissonette and co-authors<sup>18</sup> demonstrated that topiramate (a mood stabiliser acting on the GABAergic system) administration impaired the cognitive flexibility components of the ASST. Additionally, one of the key features of the ASST is represented by the fact that it leverages a series of stimuli and behavioural responses that are part of the rodent ethological repertoire.

Despite its numerous advantages, the ASST presents some core limitations. First, its execution is extremely labour intensive and requires the constant engagement of a highly-trained experimenter; second, the numerous variables involved (e.g., odours, digging media, rewards, positioning of the reward within the bowl) are likely to generate a considerable variation within and between laboratories; third, the combination of the aforementioned aspects increases the number of potential mistakes (e.g., inversion of rewarded odours); fourth, complete blinding of the experimenter is difficult to attain.

A potential solution to these limitations has been recently proposed through the design and development of a fully-automated version of the ASST<sup>29</sup>. This automated task closely mimics the WCST and the CANTAB ID/ED task, the two most widely used neuropsychological tests for the evaluation of attentional set-shifting abilities in humans. Recently, this version has been made commercially available (Operon 49500, Ugo Basile S.r.l., Gemonio, Italy<sup>30</sup>). The latter allows the execution of a modified yet analogous version of the ASST with minimal human supervision. Similar to the manual version, OPERON allows the completion of all the stages of the ASST using palatable food rewards delivered upon the execution of an operant response. OPERON also affords three dimensions (olfactory, tactile-visual, and visual) varying along several sets of stimuli as the cues to which experimental subjects have to pay attention throughout the test. Yet, these sets of stimuli are different from those offered by the manual version (in particular, for the tactile-visual dimension, texture of the conditioning area rather than digging medium). Likewise, the operant response, insertion of the nose into a nose-poke hole instead of digging, is radically different between the two versions. Ultimately, despite the theoretical similarities, the manual version of the ASST and its automated incarnation, present several discrepancies that we aim to fully unravel.

In the current study, we conducted a back-to-back comparison of the manual (ASST) and automated (OPERON) versions of this task to test whether they provide analogous results. To this aim, we approached this question through several converging avenues: first, we tested whether the two versions are equivalently sensitive to strain differences (Experiment 1); second, we assessed whether the effects of chronic stress—in the

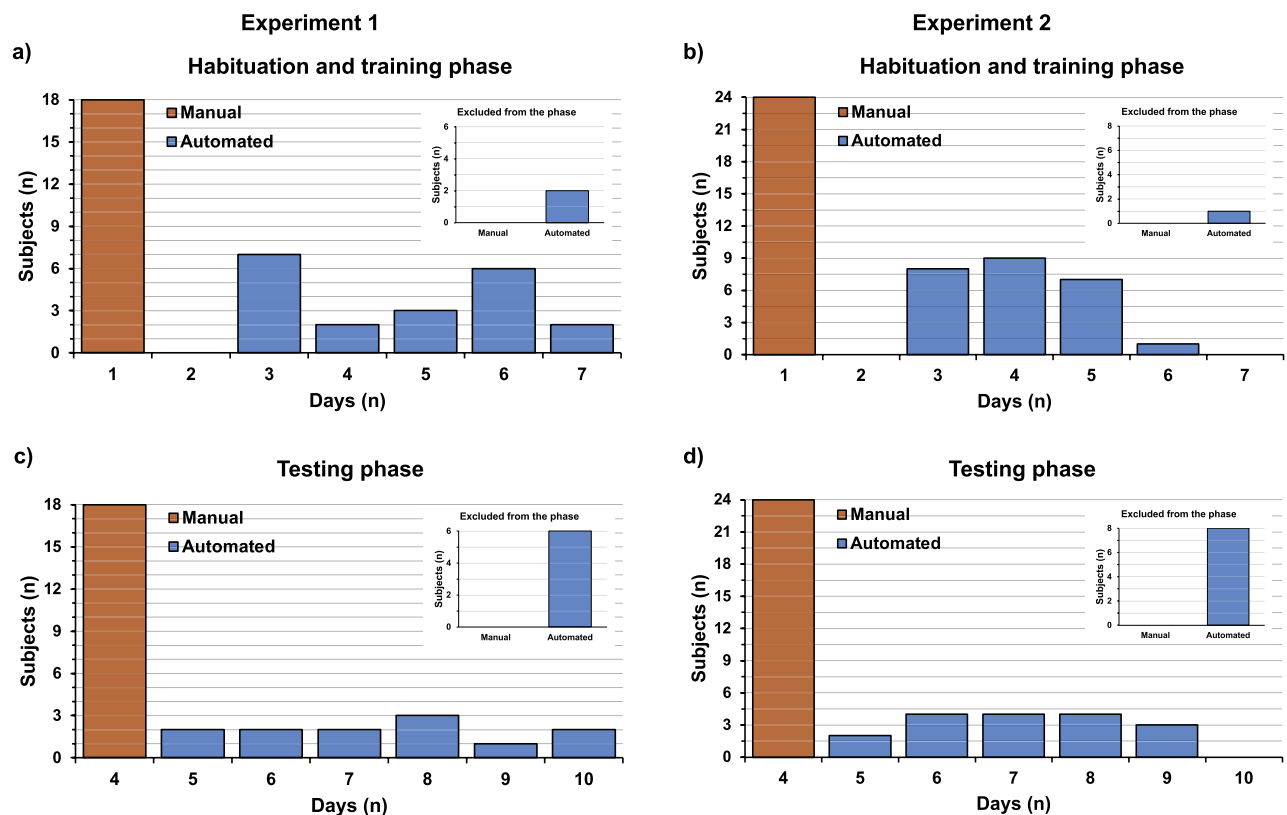
form of corticosterone (hereafter CORT) administration in the drinking water (see e.g.<sup>31</sup>)—on attentional set-shifting capabilities, previously reported to be sensitive to diverse stressors<sup>32</sup>, remain stable when tested through OPERON (Experiment 2). We thus tested, in both versions of the ASST task: (Experiment 1) independent cohorts of C57BL/6 and CD1 mice; and (Experiment 2) C57BL/6 administered CORT in the drinking water, and their controls (hereafter VEH).

## Results

The manual testing phase was performed over 4 days whilst the automated testing phase required a minimum of five and a maximum of 10 days. The habituation and training phase lasted 1 day in the manual version and a minimum of three and a maximum of seven days in the automated version. The number of days required to complete the habituation and training phase and the five stages of testing, in the manual and automated versions, are shown in Fig. 1a–d.

Regarding Experiment 1, whilst all subjects (CD1:  $n=9$ ; C57:  $n=9$ ) tested in the manual version completed the task and were therefore included in the analyses, some animals (CD1:  $n=4$ ; C57:  $n=4$ ) did not complete the OPERON as they did not reach the criterion within the established number of sessions and could not continue the test. The Fisher's exact test, performed to highlight within-strain differences in the number of animals that failed to complete the manual vs. automated ASST, did not reveal any significant differences (habituation and training phase CD1  $P=0.4737$ , C57  $P=1.000$ ; testing CD1  $P=0.2059$ , C57  $P=0.0867$ ).

In Experiment 2, all subjects tested in the manual version completed the task and were included in the analyses (VEH:  $n=12$ ; CORT:  $n=12$ ). Conversely, in the OPERON, some animals did not reach the criterion within



**Figure 1.** Number of subjects achieving the completion of the manual and automated version of the task on the specific day reported in the x-axis (main panels), or failing to complete the task (inset). **(a)** Number of mice tested in the manual (CD1:  $n=9$ ; C57:  $n=9$ ) and automated (CD1:  $n=8$ ; C57:  $n=10$ ) ASST in relation to the number of days employed to complete the habituation and training phase in Experiment 1. Inset: Mice that failed to complete the habituation and training phase of the automated ASST (CD1:  $n=2$ ; C57:  $n=0$ ). **(b)** Number of mice tested in the manual (VEH:  $n=12$ ; CORT:  $n=12$ ) and automated (VEH:  $n=12$ ; CORT:  $n=13$ ) ASST in relation to the number of days employed to complete the habituation and training phase in Experiment 2. Inset: Mice that failed to complete the habituation and training phase of the automated ASST (VEH:  $n=1$ ; CORT:  $n=0$ ). **(c)** Number of mice tested in the manual (CD1:  $n=9$ ; C57:  $n=9$ ) and automated (CD1:  $n=6$ ; C57:  $n=6$ ) ASST in relation to the number of days employed to complete the five stages of the testing phase in Experiment 1. Inset: Mice that failed to complete the testing phase of the automated ASST (CD1:  $n=2$ , 1 in CDR and 1 in EDS; C57:  $n=4$ , 1 in SD, 1 in CD, 1 in CDR, 1 in IDS). **(d)** Number of mice tested in the manual (VEH:  $n=12$ ; CORT:  $n=12$ ) and automated (VEH:  $n=10$ ; CORT:  $n=7$ ) ASST in relation to the number of days employed to complete the five stages of the testing phase in Experiment 2. Inset: Mice that failed to complete the testing phase of the automated ASST (VEH:  $n=2$ , 1 in SD and 1 in CD; CORT:  $n=6$ , 4 in SD, 1 in CD, 1 in CDR).

the established number of sessions and were excluded from the test and the analyses (VEH:  $n=3$ ; CORT:  $n=6$ ). In the absence of within-group differences in the habituation and training phase (VEH  $P=1.000$ , CORT  $P=1.000$ ), the Fisher's exact test indicated that a higher number of CORT subjects failed to complete the automated task compared to the manual ASST (VEH  $P=0.4783$ , CORT  $P=0.0149$ ).

## Separated analysis of manual and automated versions

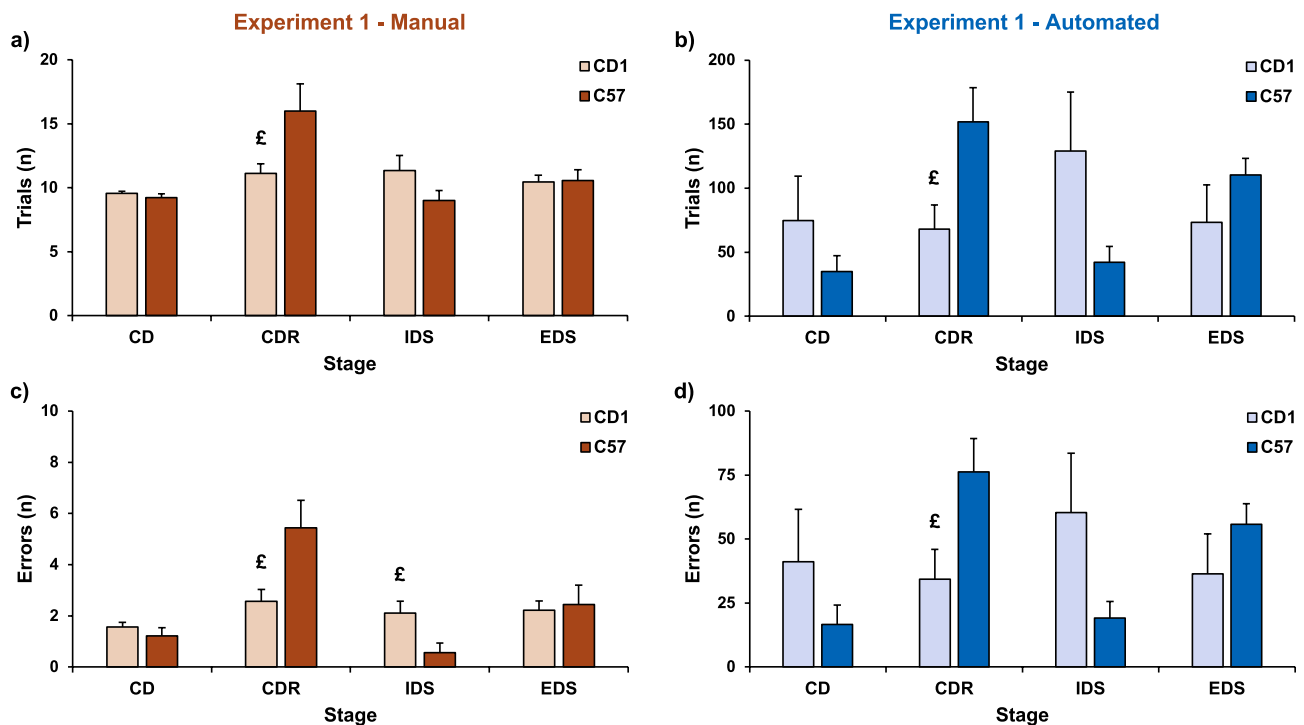
### Experiment 1

In the manual version, no differences were found between CD1 and C57BL/6 mice in the acquisition of the CD stage, which was characterised by the introduction of the irrelevant dimension (strain:  $F(1,16)=0.82$ ,  $P=0.379$  for trials, Fig. 2a;  $F(1,16)=0.82$ ,  $P=0.379$  for errors, Fig. 2c). A similar profile emerged in the CD stage of the automated version (strain:  $F(1,10)=1.20$ ,  $P=0.298$  for trials, Fig. 2b;  $F(1,10)=1.25$ ,  $P=0.289$  for errors, Fig. 2d). In the reversal of the CD stage (CDR), which required the formation of a new rule, we found a significant difference between the two strains, with decreased attentional performances in C57BL/6 compared to CD1 mice, in both the manual (strain:  $F(1,16)=4.79$ ,  $P=0.044$  for trials, Fig. 2a;  $F(1,16)=6.10$ ,  $P=0.025$  for errors, Fig. 2c) and the automated version (strain:  $F(1,10)=6.49$ ,  $P=0.029$  for trials, Fig. 2b;  $F(1,10)=5.69$ ,  $P=0.038$  for errors, Fig. 2d). Specifically, C57BL/6 mice required more trials and committed more errors than CD1 to attain the CDR, in both versions. No differences between the two strains were found in the number of trials to attain the IDS stage, either in the manual (strain:  $F(1,16)=2.72$ ,  $P=0.118$ , Fig. 2a) or in the automated (strain:  $F(1,10)=3.29$ ,  $P=0.100$ , Fig. 2b) version. However, CD1 mice committed more errors than C57BL/6 in the manual (strain:  $F(1,16)=6.94$ ,  $P=0.018$ , Fig. 2c), but not in the automated ( $F(1,10)=2.92$ ,  $P=0.118$ , Fig. 2d) version. In the manual version, no differences were found between CD1 and C57BL/6 mice in attaining the EDS stage, which required an attentional shift to the previously irrelevant dimension (strain:  $F(1,16)=0.01$ ,  $P=0.913$  for trials, Fig. 2a;  $F(1,16)=0.07$ ,  $P=0.797$  for errors; Fig. 2c). A similar profile emerged in the EDS stage of the automated version (strain:  $F(1,10)=1.34$ ,  $P=0.274$  for trials, Fig. 2b;  $F(1,10)=1.20$ ,  $P=0.299$  for errors, Fig. 2d).

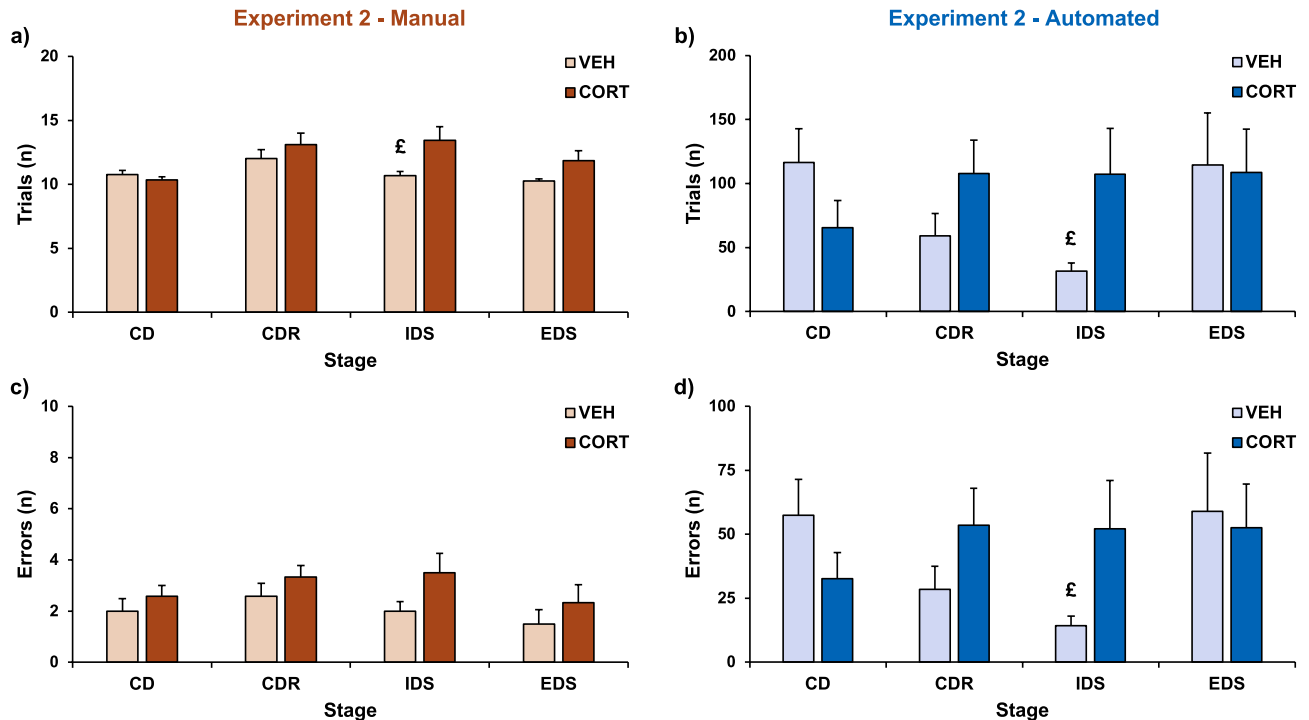
### Experiment 2

We confirmed that CORT administration significantly increased CORT concentrations in plasma. Thus, compared to VEH-treated controls, CORT-treated subjects had a substantial increase in plasma CORT concentrations (treatment:  $F(1,22)=17.93$ ,  $P<0.001$ ; CORT:  $7748.30 \pm 1695.54$  pg/ml; VEH:  $557.01 \pm 99.81$  pg/ml).

The treatment with CORT solution did not impair the performance of the mice in the CD stage both in manual (treatment:  $F(1,22)=0.92$ ,  $P=0.348$  for trials, Fig. 3a;  $F(1,22)=0.85$ ,  $P=0.367$  for errors, Fig. 3c) and



**Figure 2.** Trials to criterion and errors committed in the manual and automated version of the task in Experiment 1. (a) Mean number ( $\pm$  SEM) of trials necessary to complete each stage of the manual ASST (CD1:  $n=9$ ; C57:  $n=9$ ). (b) Mean number ( $\pm$  SEM) of trials necessary to complete each stage of the automated ASST (CD1:  $n=6$ ; C57:  $n=6$ ). (c) Mean number ( $\pm$  SEM) of errors committed to complete each stage of the manual ASST (CD1:  $n=9$ ; C57:  $n=9$ ). (d) Mean number ( $\pm$  SEM) of errors committed to complete each stage of the automated ASST (CD1:  $n=6$ ; C57:  $n=6$ ).  $^{\#}P \leq 0.05$  compared to C57 mice in post-hoc test. For data regarding the first stage of the task, which required only a simple discrimination on one stimulus dimension, see Supplementary material, paragraph 2.1.1.



**Figure 3.** Trials to criterion and errors committed in the manual and automated version of the task in Experiment 2. (a) Mean number ( $\pm$  SEM) of trials necessary to complete each stage of the manual ASST (VEH:  $n = 12$ ; CORT:  $n = 12$ ). (b) Mean number ( $\pm$  SEM) of trials necessary to complete each stage of the automated ASST (VEH:  $n = 10$ ; CORT:  $n = 7$ ). (c) Mean number ( $\pm$  SEM) of errors committed to complete each stage of the manual ASST (VEH:  $n = 12$ ; CORT:  $n = 12$ ). (d) Mean number ( $\pm$  SEM) of errors committed to complete each stage of the automated ASST (VEH:  $n = 10$ ; CORT:  $n = 7$ ).  $\epsilon P \leq 0.05$  compared to CORT mice in post-hoc test. Data regarding the first stage of the task, which required only a simple discrimination on one stimulus dimension, are reported in the Supplementary material, paragraph 2.1.2.

in automated version (treatment:  $F(1,15) = 1.95$ ,  $P = 0.183$  for trials, Fig. 3b;  $F(1,15) = 1.71$ ,  $P = 0.211$  for errors, Fig. 3d). No differences between the two groups were found in the number of trials to attain the criterion or number of errors committed in the CDR stage in manual version (treatment:  $F(1,22) = 0.89$ ,  $P = 0.357$  for trials, Fig. 3a;  $F(1,22) = 1.25$ ,  $P = 0.276$  for errors, Fig. 3c). A similar pattern emerged in the CDR stage of the automated version (treatment:  $F(1,15) = 2.59$ ,  $P = 0.129$  for trials, Fig. 3b;  $F(1,15) = 2.44$ ,  $P = 0.139$  for errors, Fig. 3d). In the IDS stage, CORT administration affected the performance of the experimental subjects, with the CORT group showing decreased attentional capabilities both in manual (treatment:  $F(1,22) = 6.03$ ,  $P = 0.022$  for trials, Fig. 3a;  $F(1,22) = 3.13$ ,  $P = 0.091$  for errors, Fig. 3c) and in automated version (treatment:  $F(1,15) = 6.22$ ,  $P = 0.025$  for trials, Fig. 3b;  $F(1,15) = 5.46$ ,  $P = 0.034$  for errors, Fig. 3d). Then, in the manual version, no differences between control and CORT subjects were found in attaining the EDS stage (treatment:  $F(1,22) = 3.94$ ,  $P = 0.06$  for trials, Fig. 3a;  $F(1,22) = 0.85$ ,  $P = 0.366$  for errors, Fig. 3c). A similar profile emerged in the EDS stage of the automated version (treatment:  $F(1,15) = 0.01$ ,  $P = 0.921$  for trials, Fig. 3b;  $F(1,15) = 0.04$ ,  $P = 0.84$  for errors, Fig. 3d).

### Combined analysis of manual and automated versions

The measures obtained in the manual and automated versions were on different scales: in Experiment 1, while in the manual version, the average number of trials and errors per stage were respectively 10.9 (range 8–30) and 2.3 (range 0–12), in the automated version, these numbers were 85.5 (range 10–295) and 42.5 (range 2–144), respectively. Experiment 2 paralleled Experiment 1 whereby, in the manual version, the average number of trials and errors were 11.9 (range 10–21) and 2.7 (range 0–8), and in the automated they were 87 (range 10–419) and 42.7 (range 2–235), respectively. Therefore, to allow the direct comparison of these measures on different scales, we devised two distinct approaches: the standardisation based on the version, intended to minimise the effect of this variable; and the calculation of an additional variable called efficiency, focusing on the proportion of errors committed (see section [Statistical analyses](#)).

#### Experiment 1

**Trials to criterion and errors committed following standardisation.** As a result of the version-based standardisation, all effects involving the version were not significant (for trials, version:  $F(1,26) = 0.05$ ,  $P = 0.832$ ); strain  $\times$  version:  $F(1,26) = 0.09$ ,  $P = 0.768$ ; stage  $\times$  version:  $F(3,78) = 0.22$ ,  $P = 0.880$ ; stage  $\times$  version  $\times$  strain:  $F(3,78) = 0.30$ ,  $P = 0.823$ ; for errors, version:  $F(1,26) = 0.06$ ,  $P = 0.813$ ; strain  $\times$  version:  $F(1,26) = 0.13$ ,  $P = 0.723$ ; stage  $\times$  version:  $F(3,78) = 0.14$ ;  $P = 0.932$ ; stage  $\times$  version  $\times$  strain:  $F(3,78) = 0.20$ ,  $P = 0.896$ ).

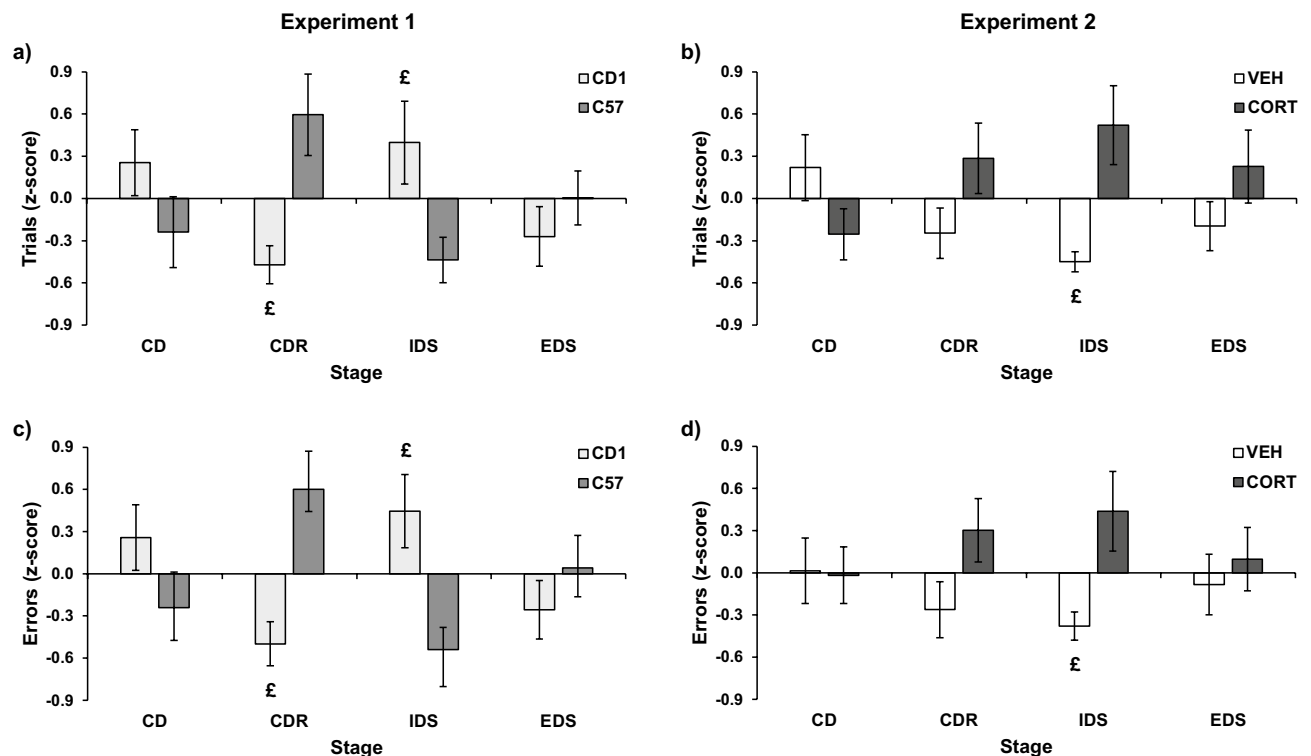
In the absence of main effects of strain ( $F(1,26) = 0.01$ ,  $P = 0.940$  for trials;  $F(1,26) = 0.002$ ,  $P = 0.962$ , for errors) and stage ( $F(3,78) = 0.18$ ,  $P = 0.910$  for trials;  $F(3,78) = 0.13$ ,  $P = 0.941$ , for errors), we found a significant stage by strain interaction for both the number of trials performed to reach the criterion and the number of errors committed ( $F(3,78) = 7.08$ ;  $P < 0.001$  and  $F(3,78) = 8.06$ ,  $P < 0.001$  respectively). This finding indicates that the differential profile of CD1 and C57BL/6 mice is maintained in both the manual and the automated task. Post-hoc analyses indicated that, compared to C57BL/6, CD1 mice performed significantly fewer trials in the CDR and significantly more trials in the IDS (Fig. 4a). A similar pattern was observed for errors (Fig. 4b).

**Efficiency.** To evaluate whether the two versions had a differential impact on the proportion of errors committed, we calculated the efficiency to complete each stage, calculated as  $[\text{Trials performed to reach the criterion} / (\text{Trials performed to reach the criterion} + \text{Errors committed})] \times 100$ . This approach allowed a direct comparison of the manual and automated versions.

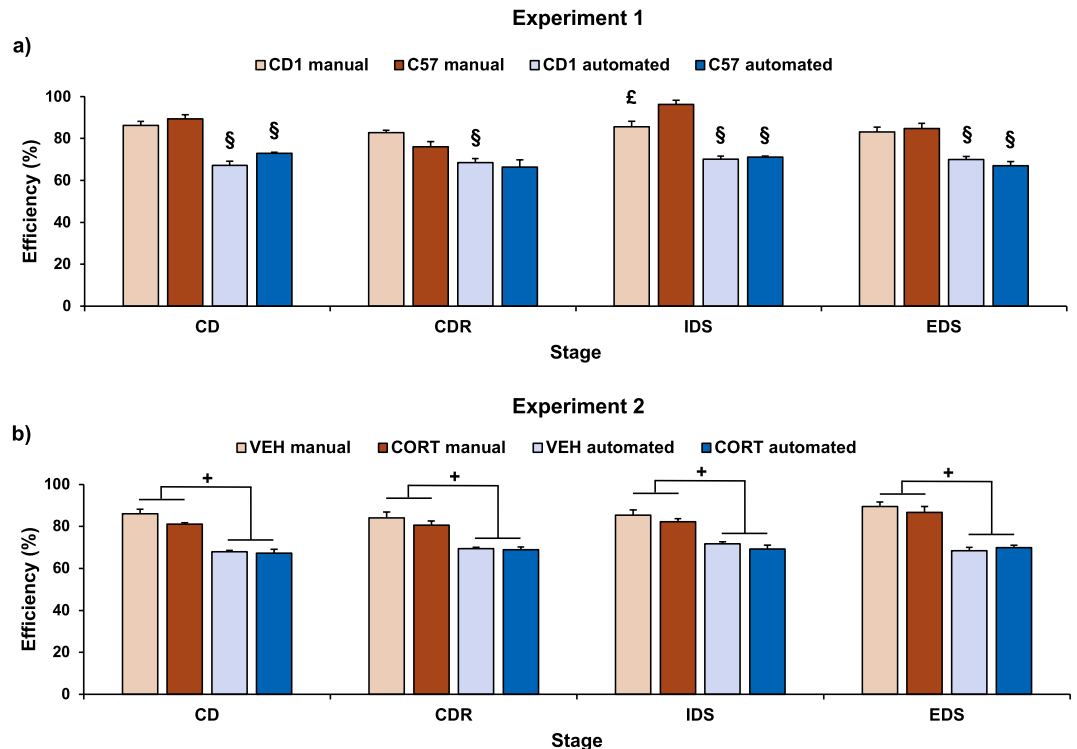
In Experiment 1, we found a main effect of the version ( $F(1,26) = 191.74$ ,  $P < 0.001$ ), with the manual task being on average significantly more efficient than the automated one. The efficiency varied from stage to stage ( $F(3,78) = 6.87$ ,  $P < 0.001$ ) depending on both the version and the strain (stage  $\times$  version:  $F(3,78) = 2.11$ ,  $P = 0.106$ ; stage  $\times$  strain:  $F(3,78) = 3.72$ ;  $P = 0.015$ ; stage  $\times$  version  $\times$  strain:  $F(3,78) = 1.80$ ,  $P = 0.154$ ). Specifically, CD1 mice performed all stages of the manual task with a significantly higher level of efficiency compared to the corresponding stages of the automated task (Fig. 5a). A similar profile was obtained for C57BL/6 mice except for the CDR stage, in which no differences of efficiency were found between the manual and the automated version. Finally, C57BL/6 mice performed the IDS stage of the manual task with a significantly higher level of efficiency compared to CD1 mice (Fig. 5a). The main effect of the strain and the strain by version interaction were not significant ( $F(1,26) = 1.22$ ,  $P = 0.280$ ;  $F(1,26) = 0.53$ ,  $P = 0.475$ , respectively).

### Experiment 2

**Trials to criterion and errors committed following standardisation.** As expected, all effects involving the version were not significant (for trials, version:  $F(1,37) = 0.03$ ,  $P = 0.863$ ); treatment  $\times$  version:  $F(1,37) = 0.29$ ,  $P = 0.592$ ; stage  $\times$  version:  $F(3,111) = 0.05$ ,  $P = 0.986$ ; stage  $\times$  version  $\times$  treatment:  $F(3,111) = 0.68$ ,  $P = 0.564$ ; for errors, version:  $F(1,37) = 0.02$ ,  $P = 0.88$ ; treatment  $\times$  version:  $F(1,37) = 0.55$ ,  $P = 0.462$ ; stage  $\times$  version:  $F(3,111) = 0.04$ ;



**Figure 4.** Trials to criterion and errors committed following a z-score normalization based on the version, intended to minimise the effect of this variable and to allow the direct comparison of measures on different scales. (a) Standardised mean number ( $\pm$  SEM) of trials necessary to complete each stage of the ASST in Experiment 1 (CD1:  $n = 15$ ; C57:  $n = 15$ ). (b) Standardised mean number ( $\pm$  SEM) of trials necessary to complete each stage of the ASST in Experiment 2 (VEH:  $n = 22$ ; CORT:  $n = 19$ ). (c) Standardised mean number ( $\pm$  SEM) of errors committed to complete each stage of the ASST in Experiment 1 (CD1:  $n = 15$ ; C57:  $n = 15$ ). (d) Standardised mean number ( $\pm$  SEM) of errors committed to complete each stage of the ASST in Experiment 2 (VEH:  $n = 22$ ; CORT:  $n = 19$ ). £ $P < 0.05$  compared to C57 mice in (a) and (c) or to CORT mice in (b) and (d) in post-hoc test.



**Figure 5.** Efficiency, intended to evaluate whether the two versions had a differential impact on the proportion of errors committed. (a) Mean efficiency ( $\pm$  SEM) to complete each stage of the manual and automated version of the ASST in Experiment 1 (manual: CD1:  $n=9$ ; C57:  $n=9$ ; automated: CD1:  $n=6$ ; C57:  $n=6$ ). (b) Mean efficiency ( $\pm$  SEM) to complete each stage of the manual and automated version of the ASST in Experiment 2 (manual: VEH:  $n=12$ ; CORT:  $n=12$ ; automated: VEH:  $n=10$ ; CORT:  $n=7$ ).  $^{\text{€}}P \leq 0.05$  compared to C57 mice;  $^{\text{§}}P \leq 0.05$  compared to the corresponding manual version in post-hoc test;  $^{\text{+}}P \leq 0.05$  main effect of the version.

$P=0.988$ ; stage  $\times$  version  $\times$  treatment:  $F(3,111)=1.02$ ,  $P=0.388$ ). In the absence of the main effect of the stage (stage:  $F(3,111)=0.05$ ,  $P=0.99$  for trials and  $F(3,111)=0.04$ ,  $P=0.99$  for errors), the minimisation of the effect of the version allowed us to show that CORT treatment affected the performance of experimental subjects both in the manual and in the automated task. Specifically, the CORT group showed reduced attentional performance compared to control subjects, requiring a higher number of trials and errors to complete the task (treatment:  $F(1,27)=6.29$ ,  $P=0.017$  for trials, Fig. 4c;  $F(1,27)=6.02$ ,  $P=0.018$  for errors, Fig. 4d). We found a significant stage by treatment interaction for trials performed to reach the criterion, with post-hoc analyses indicating that, compared to control subjects, the CORT group performed significantly more trials and committed more errors to complete the IDS stage (stage  $\times$  treatment:  $F(3,111)=3.97$ ,  $P=0.001$  for trials, Fig. 4c;  $F(3,111)=1.92$ ,  $P=0.131$  for errors, Fig. 4d).

**Efficiency.** Both VEH and CORT subjects performed all stages of the manual task with a significantly higher level of efficiency compared to the corresponding stages of the automated task, with analyses showing a significant main effect of the version ( $F(1,37)=167.365$ ,  $P<0.001$ ), regardless of treatment ( $F(1,37)=2.86$ ,  $P=0.099$ ). Besides, the efficiency did not vary depending on stage or on interaction between different factors (treatment  $\times$  version:  $F(1,37)=1.7$ ,  $P=0.201$ ; stage:  $F(3,111)=1.37$ ,  $P=0.255$ ; stage  $\times$  treatment:  $F(3,111)=0.17$ ,  $P=0.914$ ; stage  $\times$  version:  $F(3,111)=1.37$ ,  $P=0.276$ ; stage  $\times$  treatment  $\times$  version:  $F(3,111)=0.13$ ,  $P=0.942$ , Fig. 5b).

## Discussion

The core aim of our study was to challenge the validity of a newly-developed, commercially available, automated version of the ASST (Operon 49500, Ugo Basile S.r.l., Gemonio, Italy<sup>30</sup>). To this aim, we stress-tested OPERON adopting two converging experimental strategies. On the one hand, we conducted the ASST using the traditional version and its automated incarnation in C57BL/6 and CD1 male mice; on the other hand, we evaluated whether OPERON was sufficiently sensitive to capture the differential cognitive capabilities observed in an experimental model of chronic stress. We decided to validate the automated version of the ASST only in males, since previous efforts reporting strain differences in its manual version have been conducted predominantly in males. While this does not provide a valid justification for continuing to conduct these studies exclusively in males, we opted for this decision since our goal was to confirm the effectiveness of the new test in an experimental group for which the anticipated differences had already been documented<sup>33–35</sup>.

We note that this is the first instance in which an automated version of the ASST is directly compared to the manual task.

In support of the validity of OPERON, we observed that C57BL/6 mice encountered more difficulties (i.e., required more trials) than CD1 mice in the CDR, be it conducted using the manual or the automated version. Likewise, both OPERON and the manual version were sufficiently sensitive to substantiate our prediction that mice exposed to a chronic stress procedure had impaired cognitive capabilities.

Previous studies investigated strain differences in this task and consistently reported major inter-individual variations. For example, in accordance with our data, Colacicco and co-authors<sup>33</sup> observed that C57BL/6J mice exhibited impairments in selected stages of the ASST, albeit their control subjects were constituted by the inbred 129/SvEv strain and not by the outbred CD1 strain adopted herein. Accordingly, we previously reported that C57BL/6 exhibited significant impairments in the ASST task compared to ABH mice<sup>34</sup>. Strain differences in attentional capabilities between C57BL/6J and CD1 have also been reported in other behavioural paradigms. Oliver and co-authors<sup>35</sup> observed that C57BL/6J mice showed impaired attentional capabilities compared to CD1 mice in the five-choice serial reaction time task (5-CSRTT). The latter measures the ability to focus the attention on a specific information, disregarding other stimuli<sup>36</sup>. It translates to rats and mice the continuous performance tests used to study attention in human<sup>37,38</sup>, thus allowing the evaluation of attention and impulsivity<sup>39</sup>. Strain differences between C57BL/6J mice and CD1 have also been reported in working memory, another core component of EFs<sup>4,5</sup>. Specifically, Sultana and co-authors<sup>40</sup> reported that C57BL/6J mice exhibited decreased working memory compared to CD1 in the Y-maze test. These findings suggest that the impairment observed in this task in C57BL/6 mice has a remarkable degree of external validity whereby it has been reproduced in independent facilities adopting independent test strategies.

The capability of OPERON to replicate the differences observed in the manual version in two mouse strains extended to the disease model tested herein. Thus, mice chronically exposed to CORT exhibited significant impairments in IDS regardless of the ASST version adopted.

The ASST has been extensively used to reveal impairments in executive functions in several experimental models of chronic stress (for a review see<sup>32</sup>). With respect to CORT administration, Wallace and collaborators<sup>41</sup> adopted an experimental strategy analogous to that adopted herein, and, in agreement with our findings, also reported that CORT-treated subjects exhibited impaired rule abstraction (IDS). Additionally, Nikiforuk and Popik further confirmed the role of CORT in mediating attentional capabilities in experimental models of chronic stress<sup>42</sup>. Thus, they first observed that chronically stressed rats had impairments in the ASST, and then demonstrated that the inhibition of CORT synthesis had a protective effect whereby it abolished the reported impairment<sup>42</sup>. The relevance of CORT as a candidate regulator of executive functions is further supported by our study. Thus, beside the impairment in IDS, we confirmed that CORT administration successfully increased plasma CORT concentrations.

With respect to the heuristic value of the ASST, the two experiments we conducted captured between-group differences in differential stages of the task. Thus, while we observed that CD1 and C57BL/6 mice differ in the CDR stage (linked to the orbitofrontal cortex<sup>21</sup>), CORT and control mice differ in the IDS stage (linked to the cingulate cortex<sup>43</sup>). The fact that these findings were consistent regardless of the level of task automation further strengthens: (i) the sensitivity of the ASST; and (ii) the remarkable correspondence between the manual and automated versions of the task.

With respect to the plasticity and flexibility of the automated version of the ASST, we note that the OPERON software can be easily customized to manage the experimental stages. Specifically, the protocol can be adjusted relatively seamlessly to meet specific experimental requirements. For example, just as the original manual version of the ASST entailed three different perceptual dimensions (20; in rats; <sup>44</sup>in mice), so also OPERON affords three dimensions. Yet, several authors reduced the number of required dimensions to two<sup>18,25,27,33,34,45</sup>. Importantly, using two dimensions, the EDS allows to test the inability to re-engage attention towards a previously irrelevant dimension (i.e., “learned irrelevance”), whilst with three dimensions it is possible to use the EDS to test for the inability to release attention from a relevant perception dimension (i.e., perseveration or “stuck-in-set”<sup>30,46</sup>). Such downscaling can be, and has been herein, easily attained in OPERON. Likewise, the number of stages adopted in literature varies between experiments. To account for this, OPERON offers a basic series of nine stages (as in the CANTAB), which can be easily reduced to meet specific experimental needs (e.g., SD, CD, CDR, IDS, EDS as in<sup>27,28,34</sup>).

The flexibility of OPERON allowed us to implement some task modifications that promoted the procedural and theoretical similarity with the manual version. Our modifications aimed at replicating in OPERON the performance-based nature of the manual version. Specifically, in the manual version, mice are allowed to continue performing the task as long as they are motivated to. Conversely, in the original version of OPERON<sup>30</sup>, mice were tested on a fixed schedule that would end upon attainment of a predetermined number of trials. We believe that, beside promoting the similarity between the two versions of the task, this expedient has the advantage of meeting individual needs and predispositions.

In the comparison of the two tasks, we observed that the efficiency of the manual version is remarkably higher than that of the automated one. Thus, the ratio between errors and trials to criterion is much lower in the traditional version, indicating that mice learn to discriminate the incorrect from the correct stimulus at a much faster pace. This result is likely related to the fact that the manual version has a much higher ethological validity compared to OPERON, whereby it rests upon a species-specific highly meaningful behavioural response (digging). This feature translates into the fact that the manual ASST barely requires any training and entails only a few days of testing. Whilst this aspect certainly results in a quicker completion of the task, it nonetheless may trigger potential floor effects, ultimately hampering the observation of subtle between-group differences.

One last aspect that warrants consideration relates to the number of subjects used in both versions of the task. In the present study, while all subjects tested in the manual version completed the task, approximately 40%



of subjects tested in OPERON had to be discarded since they failed to reach the criterion within the established number of sessions. It is important to emphasize that a certain experimental dropout is the norm rather than an exception in operant tasks, e.g.<sup>47</sup>. We note, however, that the reported dropout did not hinder the quality of the observed findings. Albeit reduced in number, the animals tested in the automated version yielded experimental data that were analogous or even better in quality to those afforded by mice tested in the manual task. Specifically, C57BL/6 mice tested in the automated version required a higher number of trials to attain the EDS compared to the IDS. While this phenotype was predicted based on available literature (e.g.<sup>43</sup>), it was not observed in the manual version. Interestingly, similar patterns of performance were reported in humans, monkeys, rats, and mice, with initial difficulties in achieving the CDR and improved performance in the IDS compared with EDS<sup>20,22,29,44,47,48</sup>.

Ultimately, the detailed back-to-back comparison conducted in this study allows precise considerations regarding methodological and practical aspects potentially guiding scientists interested in the ASST. From a purely methodological perspective, compared to the manual version, OPERON requires a much lower level of training in behavioural neuroscience and affords a much higher degree of standardisation and automation. The latter also protects from potential mistakes (e.g., inversion of rewarded stimuli), which may occur in the light of the labour-intensive nature of the task and its associated tiredness. Additionally, OPERON guarantees by definition full experimenter blinding whereby it is fully automated and does not allow any room for the experimenter to unconsciously influence the response of the tested individual. For example, in our study, blinding was not possible since the experimental groups differed for the coat colour. In spite of these advantages, OPERON does not perform on par with the manual version concerning other relevant aspects. For example, the behavioural response required to obtain the reward in the manual version is much more ethologically relevant than that required in OPERON; this likely translates in the higher efficiency observed in the manual version, and in its largely reduced duration (two- to three-fold) compared to OPERON. We cannot exclude that this aspect also contributes to the lack of dropouts observed in the manual ASST. Finally, the economic dimension of the two versions of the ASST warrants consideration. While the initial investment is certainly in favour of the manual version (few hundred euros versus several thousands), an overall budget estimation cannot neglect the human costs associated with the conduction of this test. While the manual version requires a full-time commitment of a highly trained operator throughout the entire study, OPERON demands very minimal experimental supervision.

In conclusion, we show that, despite minor caveats, the automated version of the ASST bestows results analogous to those afforded by the manual version, while limiting manpower and sources of confound.

## Methods

### Ethics statement

All experimental procedures were approved by Institutional Animal Survey Board on behalf of the Italian Ministry of Health (licence n. 729/2020-PR to SM) and performed in full accordance with the Directive 2010/63/EU on the protection of animals used for scientific purposes and Italian law (Legislative Decree 26/2014). All sections of this report adhere to the ARRIVE Guidelines for reporting animal research. A completed ARRIVE Essential 10 Checklist is included as supplemental material.

### Animals and experimental design

#### Experiment 1

Experimental subjects were 19 adult male C57BL/6N mice and 19 adult male CD1 mice provided by Charles River Laboratories s.r.l. (Calco, Italy). Upon arrival C57BL/6N and CD1 mice, weighting approx. 20 g and 24 g respectively, were housed in pairs in polycarbonate cages (33.0 × 13.0 × 14.0 cm) equipped with metal tops. All animals had access to sawdust bedding, environmental enrichment in the form of shelter material (Nestlets<sup>®</sup>), and ad libitum water and food (Mucedola s.r.l., Settimo Milanese, Italy) until the beginning of the experiment (for details on the food restriction schedule see next paragraph). Animals were identified through ear-clipping and housed in an air-conditioned room (temperature 24 ± 1 °C, relative humidity 40 ± 5%), on a 12-h reversed light–dark cycle (lights on at 20:00). Within each strain, mice were randomly assigned to either the manual or the automated version of the ASST. Mice were tested daily, 7 days a week (except for the habituation and training phase of the automated ASST with 6 days of testing per week), under dim-light conditions. Behavioural tests were performed in an experimental room adjacent to the housing room, minimising gradients in light, temperature, sound, and other environmental conditions. To minimise potential confounders, all experimental subjects were housed in the same room; additionally, the position of the cages within the rack was randomly assigned. Animals were brought to the experimental room inside their home-cages, which were covered with a thick cloth. Testing started around 09:00 (i.e., 1 h after the light–dark shift) and ended no later than 19:00.

Regarding group allocation, conduct of the experiment, and outcome assessment, in the manual version blinding was not possible since (i) the experimental groups differed for the coat colour and (ii) data were manually scored by the experimenter. By contrast, the automated version guaranteed full experimenter blinding in the conduct of the experiment and outcome assessment.

#### Experiment 2

Experimental subjects were 50 adult male C57BL/6J OlaHsd (approx. 20 g), provided by ENVIGO RMS s.r.l. (San Pietro al Natisone, Italy). Upon arrival, mice were housed as described in Experiment 1. After 10 days of habituation, half of the mice were randomly assigned to a CORT treatment while the other half to VEH using online random number generator. CORT (Sigma-Aldrich, Missouri, USA) treated mice received CORT dissolved in their drinking water at a dose of 0.035 mg/ml, starting from 3 weeks before the beginning and throughout the duration of the experiment. The single mouse constituted the experimental unit.

The average liquid intake was  $5.78 \pm 0.16$  ml per day. In consequence of the ad libitum access to the CORT solution, mice received on average  $8.60 \pm 0.33$  mg/kg/day of CORT. Details regarding the preparation of the CORT and VEH solutions are reported in the Supplementary material (paragraph 4.2.2.). Testing of mice was performed as described in Experiment 1. At the end of the experiment (after 8 weeks of CORT treatment), blood samples were collected from CORT ( $n = 12$ ) and VEH ( $n = 12$ ) mice tested in the manual ASST to evaluate basal CORT concentrations in plasma (see Supplementary material, paragraph 4.2.2., for details).

For the manual version, as already mentioned, in the conduct of the experiment and outcome assessment blinding was not possible since. By contrast, the automated version guaranteed full experimenter blinding in the conduct of the experiment and outcome assessment. In addition, for both versions, the experimenter was not aware of group allocation.

### Food restriction schedule

In both Experiment 1 and 2, the food restriction, applied to increase animals' motivation to perform the task, aimed at maintaining the animals at 85–90% of their free feeding bodyweight. Before food restriction, the average body weight of experimental subjects in Experiment 1 was:  $25.8 \pm 0.4$  (C57BL/6N) and  $30.2 \pm 0.5$  (CD1). The average body weight of experimental subjects in Experiment 2 was:  $26.7 \pm 0.6$  (VEH) and  $26.4 \pm 0.4$  (CORT). Food restriction was applied, 7 days a week by giving mice access to food for a limited time after completing the daily session. Specifically, mice were daily weighed and the time available for access to food (usually in the range between 2 and 4 h) was individually adjusted in order to allow the subject's bodyweight to remain constant. To this aim, the order of testing was daily scheduled based on bodyweight percentages (from the lower to the higher), also to guarantee an extended access to food (e.g. 4 h) when needed. Mice were also daily inspected to identify potential signs of poor wellbeing (e.g., locomotion, fur condition). Mice were food-restricted from 3 days before the beginning of the habituation and training phase and throughout the testing phase.

### Manual version of the attentional set-shifting task (ASST)

In both Experiment 1 and 2, the manual test was conducted in a custom-made opaque PVC U-maze ( $45 \times 30 \times 15$  cm) with one starting compartment ( $30 \times 30$  cm) connected, through a sliding door, to two identical choice compartments ( $15 \times 15$  cm; Fig. 6a). A metal bowl (4.0 cm high, 7.0 cm top diameter, 4.0 cm bottom diameter) was placed in each of the two identical compartments and a small piece of cereal (1/4 honey Cheerios®, Nestlé, Vevey, Switzerland), covered with 2.0 cm layer of digging medium, was used as a reward (as in<sup>26</sup>). Under the wire-mesh floor of the maze we placed a layer of sawdust bedding from the home-cage of the tested subject to provide a more familiar environment and to reduce the stress due to exposure to a novel environment. Between sessions, the apparatus was cleaned with 30% ethanol/water solution to remove odour cues.

#### *Habituation and training phase*

The training phase was performed on the first day and consisted of nine trials, preceded by 5 min of habituation to the apparatus. This phase aimed at training mice to reliably dig into the bowls to obtain the reward. During habituation, the sliding door was removed, both bowls were empty, and mice were free to explore the whole apparatus. During training, each trial started with the mouse positioned in the starting compartment and the sliding door closed. When the mouse was facing the wall of the apparatus opposite to the door, the latter was manually raised by the experimenter, allowing contemporary access to both choice compartments. During training, both bowls were baited, and mice had to retrieve and consume both rewards to complete the trial. When the mouse spontaneously returned to the starting compartment after having retrieved the second reward, the trial was terminated by closing the sliding door. The nine trials of the training phase were arranged as follows: rewards placed at the bottom of the empty bowls in trials 1–3; rewards placed on top of the bowls filled with the same type of sawdust used as bedding material in trials 4–6; rewards covered by a 2.0 cm layer of bedding material in trials 7–9.

#### *Testing phase*

The testing phase consisted of five consecutive stages, with different discriminations involving stimuli belonging to two dimensions (i.e., olfactory and tactile-visual). The stimuli presented in each stage (i.e., odours and digging media) are shown in Fig. 7a. In the first stage (simple discrimination; SD), the animal was presented with only the olfactory dimension, having to perform a simple discrimination between two different stimuli (i.e., two odours). In the second stage (compound discrimination; CD), we introduced the tactile-visual dimension with two additional stimuli (i.e., two digging media); however, the relevant dimension (i.e., olfactory dimension) and the rewarded stimulus remained consistent with the SD. In the third stage (compound discrimination reversal; CDR), the four stimuli were maintained from the CD; the irrelevant dimension in the previous stage (i.e., tactile-visual dimension) continued to be irrelevant but the rewarded stimulus was now the one that had been previously incorrect. For the fourth (intra-dimensional shift; IDS) and fifth (extra-dimensional shift; EDS) stages, we introduced four novel stimuli in both dimensions (i.e., two odours and two digging media). Whilst in the IDS the relevant dimension remained the olfactory one, in the EDS subjects had to shift their attention to the previously irrelevant dimension (i.e., tactile-visual dimension).

At the beginning of each testing trial the mouse was positioned in the starting compartment and the sliding door was closed. As soon as the mouse happened to be far from the door and facing the opposite wall, the experimenter manually raised the sliding door to give the subject access to the two choice compartments. During testing, only one bowl was baited. Choice was defined as the first evident digging in either the baited or the unbaited bowl. Each stage began with four exploratory trials<sup>20,49</sup> during which the door remained open even if the first choice was incorrect, and the mouse was allowed to investigate and collect the reward from the opposite



**Figure 6.** Setting for the manual and automated attentional set-shifting task (ASST) in mice. (a) Photographs of the apparatus for the manual task taken from above (on the left) and of odours and digging media used as stimuli (on the right). (b) Photograph of the apparatus (Operon 49500, Ugo Basile S.r.l., Gemonio, Italy) for the automated task (in the middle), with the odour-delivery unit (on the left) and the pump for the delivery/aspiration system (on the right). In manual ASST, we mixed small amounts of powdered aromatic herbs/spices with the digging media. In automated ASST, odours were prepared by diluting liquid flavours employed in the confectionery industry (5% in mineral oil).

bowl. The outcome of these initial trials was included in the presented results (as in<sup>27</sup>). In subsequent trials, (i) if the subject started to dig in the unbaited bowl, an error was recorded, and the trial was terminated by closing the door and leaving the mouse within the choice compartment for 1 min (punishing timeout); (ii) if the subject started to dig in the baited bowl, the trial was terminated by closing the door and allowing the mouse 1 min to consume the reward within the choice compartment. In both correct and incorrect trials, after the 1-min timeout, the sliding door was opened, and the subject spontaneously returned to the starting compartment.

In order to complete each stage, mice had to reach a criterion of eight correct discriminations out of ten consecutive trials. Each day testing continued until the mouse was willing to work (usually about 60 min); specifically, a 10-min cutoff was set for each trial, allowing the experimenter to terminate testing when the mouse did not express any choice in the 10 min following the beginning of a trial.

#### Automated version of the attentional set-shifting task (ASST)

The test was conducted using the upgraded version of the OPERON task, originally developed by Scheggia and Papaleo<sup>29,50</sup>, that is now commercially available (Operon 49500, Ugo Basile S.r.l., Gemonio, Italy). For a thorough description of the apparatus, the reader is referred to Scarsi and co-authors<sup>30</sup>, where the commercially available version was firstly used.

Briefly, the apparatus consisted of a double-chambered cage, with transparent walls and metal floor, divided by an automated metal sliding door placed in the middle of the cage, and with an operant stimulation wall mounted on each side (Fig. 6b).

Each operant wall included automated stimulators involving stimuli belonging to three dimensions (i.e., olfactory, tactile-visual, visual). Specifically, each stimulation wall was equipped with an external pellet dispenser, a central pellet receptacle where precision pellets were delivered (when the correct choice was made), a single

a)

MANUAL VERSION					
Stage	Relevant dimension	Rewarded stimulus	Discrimination 1	Discrimination 2	Day
SD	olfactory	cinnamon	cinnamon vs. sage (in bedding material)	-	2
CD	olfactory	cinnamon	sawdust + cinnamon vs. cotton + sage	sawdust + sage vs. cotton + cinnamon	2
CDR	olfactory	sage	sawdust + cinnamon vs. cotton + sage	sawdust + sage vs. cotton + cinnamon	3
IDS	olfactory	ginger	crepe-paper + ginger vs. confetti + cumin	crepe-paper + cumin vs. confetti + ginger	4
EDS	tactile-visual	polystyrene	polystyrene + cloves vs. tissue-paper + rosemary	polystyrene + rosemary vs. tissue-paper + cloves	5

b)

AUTOMATED VERSION					
Stage	Relevant dimension	Rewarded stimulus	Discrimination 1	Discrimination 2	Day
SD	olfactory	cinnamon	cinnamon vs. sage	-	4-7
CD	olfactory	cinnamon	sandpaper + cinnamon vs. blackboard + sage	sandpaper + sage vs. blackboard + cinnamon	5-10
CDR	olfactory	sage	sandpaper + cinnamon vs. blackboard + sage	sandpaper + sage vs. blackboard + cinnamon	6-12
IDS	olfactory	lavender	rough cardboard + lavender vs. smooth cardboard + oregano	rough cardboard + oregano vs. smooth cardboard + lavender	9-13
EDS	tactile-visual	holed plastic foam	holed plastic foam + vanilla vs. aluminium + green tea	holed plastic foam + green tea vs. aluminium + vanilla	10-15

**Figure 7.** Schematic representation of the five stages of the (a) manual and (b) automated ASST testing phase (first column), with the dimension that was relevant (second column) and the stimulus that was rewarded (third and seventh columns) in each stage (brown and blue are used for the manual and automated versions respectively). In each stage, mice were presented with either Discrimination 1 (fourth and eighth column) or Discrimination 2 (fifth and ninth columns), according to a pseudo-random sequence. The side where each pair of stimuli was presented (left and right bowls in the manual task, left and right nose-poke holes in the automated task) were counterbalanced through a pseudo-random sequence. In the sixth and tenth columns we reported the range of days in which each stage of the task was completed.

house light placed in the top middle of the wall (signalling the length of the punishing timeout when the incorrect choice was made), two nose-poke holes (one of each side of the pellet receptacle), two programmable arrays of coloured led lights (providing the stimuli belonging to the visual dimension) placed over each nose-poke hole, and two texture floors (providing the stimuli belonging to the tactile-visual dimension) placed under each nose-poke hole (Fig. 8). The latter were part of the revolving floor system for automated tactile stimulation, which was mounted under the floor in front of each stimulation wall.



**Figure 8.** Operon apparatus. On the right, side view of the double-chambered cage (Operon 49500, Ugo Basile S.r.l., Gemonio, Italy), divided by the central sliding door, with an operant stimulation wall mounted on each side. On the left, particular of the stimulation wall (front view) highlighting the position of the following elements: external pellet dispenser (1); house light placed in the top middle of the wall, signalling the length of the punishing timeout (2); central pellet receptacle where precision pellets were delivered (3); programmable arrays of led lights, providing the stimuli belonging to the visual dimension, placed over each nose-poke hole (4); nose-poke holes equipped with the odour delivery/aspiration system, providing the stimuli belonging to the olfactory dimension (5); texture floors, providing the stimuli belonging to the tactile-visual dimension, placed under each nose-poke hole (6).

The nose-poke holes were equipped with an odour delivery/aspiration system (providing the stimuli belonging to the olfactory dimension) and connected to the odour-delivery unit, which provided automated odour stimulation via different odours (up to ten) independently in the two chambers. The odour-delivery unit was composed of ten flasks (one for each odour) with individual pressure control. Importantly, the delivery/aspiration system allowed to both deliver the odour inside the nose-poke hole and to extract the air from it after the stimulation, cleaning the hole from the odour previously used before the initiation of the subsequent trial.

The setting allowed the subject to interact with the pair of stimuli (i.e., odours and textures; the visual dimension, involving the coloured led lights, was not used in the present experiment) presented on the two sides of each operant wall, and to express its choice by inserting the nose in one of the two nose-poke holes. A 14 mg non-flavoured precision pellet (5TUL purified rodent tablets, catalog n. 1811215; TestDiet, St. Louis, MO 63144, USA), delivered in the pellet receptacle, was used as a reward.

The two chambers allowed for continuous trial repetition: at the beginning of the session the mouse was confined by the sliding door in the first chamber; after having interacted with the stimulation wall inside this chamber, the sliding door lowered, and the mouse shuttled to the opposite chamber. After having expressed its choice between the two nose-poke holes of the stimulation wall inside the second chamber, the mouse shuttled back to the first chamber where new pairs of stimuli (i.e., odours and textures) had been automatically arranged meanwhile. During both the training and the testing phase, the total number of trials that could be performed within each session was not predetermined, as mice were free to express nose-poking at their own, individually variable rate.

Infrared photo-beams, placed inside the nose-poke holes, the pellet receptacle and in proximity to the sliding door, were used to detect the activity of the animals (i.e., nose-poking, eating the rewards, shuttling to and fro). Following the delivery of the reward from the pellet dispenser, only the photo-beam inside the pellet receptacle was active. Hence, additional nose-poking was without any consequences until the interruption of the pellet receptacle photo-beam. After the reward had been eaten (in correct trials) or after the 5-s punishing timeout (in incorrect trials), the sliding door lowered, and the subject spontaneously moved to the opposite chamber to start the next trial. As soon as the photo-beam of the new chamber was interrupted, the sliding door automatically rose, confining the mouse in this chamber.

A dedicated version of the ANY-maze software (Stoelting Co., Wood Dale, IL, USA) automatically managed all OPERON cues, tools and sensors following a precompiled protocol and collected the data. Between sessions, the apparatus was cleaned with 30% ethanol/water solution to remove odour cues.

#### *Habituation and training phase*

The habituation was conducted on the first 2 days and consisted of two parts, each having a fixed duration of 40 min.

During the first part of the habituation, performed on the first day, lights were turned off, no odours were delivered inside the nose-poke holes, the texture floors were neutral (i.e., made of the same material of the apparatus floor), and the sliding door remained lowered. Inserting the nose in any of the four available nose-poke holes resulted in the delivery of the reward in the pellet receptacle of the corresponding operant wall. Until the interruption of the pellet receptacle photo-beam, additional nose-poking did not result in the delivery of additional pellets.

The second part of the habituation, performed on the second day, was similar to the first one, except for the sliding door that was activated. At the beginning of this second part of the habituation, the animal was confined by the sliding door in the first chamber, the door was then automatically lowered, allowing the animal to move to the second chamber, thus starting the first trial; as soon as the animal had entered this chamber, the sliding door automatically rose. Nose-poking in either of the two holes of the operant wall within the chamber resulted in the delivery of the reward in the pellet receptacle. Following the interruption of the pellet receptacle photo-beam, the sliding door lowered, and the animal had to move back to the first chamber to start the subsequent trial.

The training phase was conducted from the third day onwards and consisted of a variable number of daily sessions (up to five) depending on the animal performance (see below). Each daily session had a maximum duration of 40 min. In this phase, the animal was presented with the olfactory dimension, having to perform a simple discrimination between two different stimuli (i.e., two odours). Hence, only one of the two nose-poke holes was associated with the reward. The correct stimulus was randomly presented in either the right or the left nose-poke hole of the operant wall. The training phase began with ten exploratory trials during which the mouse was allowed to nose-poke in the other hole and collect the reward even if the first choice was incorrect. In subsequent trials, (i) if the subject nose-poked in the incorrect hole, an error was recorded and the trial was terminated by turning on the house light to signal the punishing timeout whilst all photo-beams were inactivated; (ii) if the subject nose-poked in the correct hole, the reward was delivered in the pellet receptacle. After the 5-s punishing timeout (in incorrect trials) or after the reward had been eaten (in correct trials), the sliding door lowered, and the subject spontaneously moved to the opposite chamber to start the next trial. As soon as the photo-beam of the new chamber was interrupted, the sliding door automatically rose confining the mouse in this chamber. In order to complete the training phase, mice had to reach a criterion of eight correct discriminations out of ten consecutive trials (the ten exploratory trials were included in the scoring). Animals that failed to reach this criterion during the available five daily sessions were excluded from the testing phase.

#### *Testing phase*

As in the manual version of the task, the testing phase consisted of five consecutive stages, with different discriminations involving stimuli belonging to two dimensions (i.e., olfactory and tactile-visual). The stimuli (i.e., odours and texture floors) presented in each stage are shown in Fig. 7b. For a description of the five stages the

reader is referred to the corresponding paragraph in the section “[Manual version of the attentional set-shifting task \(ASST\)](#)”, as the only difference was the use of texture floors in place of the digging media. Each stage of the testing phase consisted of a variable number of daily sessions (up to three for SD, CD and IDS, and up to five for CDR and EDS) depending on the animal performance.

As for the functioning of the apparatus (e.g., nose-poke holes, sliding door, photo-beams, house light, etc.), the testing phase was similar to the training phase, except for the texture floors that, from the CD stage onwards, were not neutral anymore. The number of exploratory trials at the beginning of each stage was reduced from ten to four. The maximum duration of the daily session was increased from 40 to 60 min. As in the manual version, we wanted testing to continue each day until the mouse was willing to work; to this aim, when the mouse did not express any choice while being confined within a chamber for a period of 10 min, the sliding door automatically lowered allowing the mouse to shuttle to the opposite chamber and start a new trial. However, if a similar occurrence happened a second time within the same daily session, testing was automatically terminated. The testing session also ceased if the animal failed to shuttle to the opposite chamber once the sliding door had been lowered for a period of 5 min.

As in the manual version, in order to complete each stage, mice had to reach a criterion of eight correct discriminations out of ten consecutive trials (the four exploratory trials were included in the scoring). Animals that failed to reach this criterion during the daily sessions available for each stage were excluded from the testing phase.

### Statistical analyses

For all parameters investigated, the sample size was estimated based on a sample size calculation using data from previous experiments, employing the free software G\*Power 3.1, considering  $\alpha = 0.05$ ,  $1 - \beta = 0.80$  and a large effect size (Cohen's  $f = 0.75$ ).

In both the manual and the automated version, we scored the total number of trials necessary to complete each stage of the task, which provides a measure of cognitive flexibility. Errors to criterion were also counted to have an additional measure of individual performance.

Firstly, we conducted separated analyses of the manual and the automated versions to investigate between-strain differences within each stage by performing stage by stage analyses of variance (ANOVA) with strain (2 levels: CD1 vs. C57BL/6 in Experiment 1) or group (2 levels: VEH vs. CORT in Experiment 2) as between-subjects factor.

Secondly, we conducted combined analysis of the manual and the automated versions. As measures (i.e., number of trials, number of errors) obtained in the two versions of the task were on different scales (see the “[Results](#)” section for additional details), and therefore could not be directly compared, we devised two distinct approaches: (iii) the z-score normalization of both trials and errors based on the version, intended to minimise the effect of this variable; (iv) the calculation of a variable called efficiency, which allowed to evaluate whether the two versions had a differential impact on the proportion of errors committed.

For the z-score normalization we used the following formula:  $z_i = (x_i - \bar{x})/s$ , where  $z_i$  is the resulting z score,  $x_i$  the value to be standardised,  $\bar{x}$  the sample mean (considering the two versions separately) and  $s$  the sample standard deviation (considering the two versions separately). The efficiency to complete each stage was calculated using the following formula:  $[\text{Trials performed to reach the criterion} / (\text{Trials performed to reach the criterion} + \text{Errors committed})] \times 100$ .

The number of trials and errors following standardisation and the efficiency were analysed by performing a repeated measures ANOVA with strain (2 levels: CD1 vs. C57BL/6 in Experiment 1) or group (2 levels: VEH vs. CORT in Experiment 2) and version (2 levels: manual, automated) as between-subjects factors and stage (4 levels: CD, CDR, IDS, EDS) as within-subject factor. For multiple post-hoc comparisons we opted for the Tukey test as it is a conservative post-hoc approach (i.e. the calculation of the minimal significant difference is protected against false positive findings).

Finally, Fisher's exact test was performed to evaluate potential within-group differences in the number of animals that failed to complete the manual vs. automated ASST.

All statistical analyses were conducted using the software StatView 5.0 (Abacus Concepts, USA). Data are always expressed as mean  $\pm$  SEM. Significance level was set at  $P \leq 0.05$ .

### Data availability

Raw data are available, to interested scientists, upon request to the corresponding author.

Received: 31 May 2023; Accepted: 26 September 2023

Published online: 06 October 2023

### References

1. Diamond, A. Executive functions. *Annu. Rev. Psychol.* **64**, 135–168 (2013).
2. Duncan, J. & Owen, A. M. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* **23**, 475–483 (2000).
3. Aron, A. R. The neural basis of inhibition in cognitive control. *Neuroscientist* **13**, 214–228 (2007).
4. Baddeley, A. D. & Hitch, G. Working memory. *Psychol. Learn. Motiv.* **8**, 47–89 (1974).
5. Cowan, N. An embedded process model of working memory. In *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (eds Miyake, A. & Shah, P.) 62–101 (Cambridge University Press, 1999).
6. Dajani, D. R. & Uddin, L. Q. Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends Neurosci.* **38**, 571–578 (2015).
7. Pecora, G. *et al.* Executive function. In *Encyclopedia of Animal Cognition and Behavior* (eds Vonk, J. & Shackelford, T. K.) 1–5 (Springer, 2017).

8. Nyberg, L. Cognitive control in the prefrontal cortex: A central or distributed executive?. *Scand. J. Psychol.* **59**, 62–65 (2018).
9. Powell, E. M. & Ragozzino, M. E. Cognitive flexibility: Development, disease and treatment. *Neuroscience* **345**, 1–2 (2017).
10. Ruiz-Gutiérrez, J., Arias-Sánchez, S. & Martín-Monzón, I. Neuropsychology of executive functions in patients with focal lesion in the prefrontal cortex: A systematic review. *Brain Cogn.* **146**, 105633 (2020).
11. Seniow, J. Executive dysfunctions and frontal syndromes. *Front. Neurol. Neurosci.* **30**, 50–53 (2012).
12. Steinke, A. & Kopp, B. Toward a computational neuropsychology of cognitive flexibility. *Brain Sci.* **10**, 1000 (2020).
13. Grant, D. A. & Berg, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. Exp. Psychol.* **38**, 404–411 (1948).
14. Reber, J. & Tranel, D. Frontal lobe syndromes. *Handb. Clin. Neurol.* **163**, 147–164 (2019).
15. Barnett, J. H. *et al.* Assessing cognitive function in clinical trials of schizophrenia. *Neurosci. Biobehav. Rev.* **34**, 1161–1177 (2010).
16. Keeler, J. F. & Robbins, T. W. Translating cognition from animals to humans. *Biochem. Pharmacol.* **81**, 1356–1366 (2011).
17. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* Vol. 3 (American Psychiatric Association, 2013).
18. Bissonette, G. B., Lande, M. D., Martins, G. J. & Powell, E. M. Versatility of the mouse reversal/set-shifting test: Effects of topiramate and sex. *Physiol. Behav.* **107**, 781–786 (2012).
19. Janitzky, K. *et al.* Optogenetic silencing of locus coeruleus activity in mice impairs cognitive flexibility in an attentional set-shifting task. *Front. Behav. Neurosci.* **9**, 286 (2015).
20. Birrell, J. M. & Brown, V. J. Medial frontal cortex mediates perceptual attentional set shifting in the rat. *J. Neurosci.* **20**, 4320–4324 (2000).
21. McAlonan, K. & Brown, V. Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behav. Brain Res.* **146**, 97–103 (2003).
22. Scheggia, D. *et al.* Variations in Dysbindin-1 are associated with cognitive response to antipsychotic drug treatment. *Nat. Commun.* **9**, 2265 (2018).
23. Ciampoli, M., Scheggia, D. & Papaleo, F. Automatic intra-/extra-dimensional attentional set-shifting task in adolescent mice. *Front. Behav. Neurosci.* **15**, 704684 (2021).
24. Macri, S. *et al.* Resilience and vulnerability are dose-dependently related to neonatal stressors in mice. *Horm. Behav.* **56**, 391–398 (2009).
25. Zhang, F., Yuan, S., Shao, F. & Wang, W. Adolescent social defeat induced alterations in social behavior and cognitive flexibility in adult mice: Effects of developmental stage and social condition. *Front. Behav. Neurosci.* **10**, 149 (2016).
26. Tait, D. S., Bowman, E. M., Neuwirth, L. S. & Brown, V. J. Assessment of intradimensional/extradimensional attentional set-shifting in rats. *Neurosci. Biobehav. Rev.* **89**, 72–84 (2018).
27. Hauser, J. *et al.* Sialylated human milk oligosaccharides program cognitive development through a non-genomic transmission mode. *Mol. Psychiatry* **26**, 2854–2871 (2021).
28. Zoratto, F., Franchi, F., Macri, S. & Laviola, G. Methylphenidate administration promotes sociability and reduces aggression in a mouse model of callousness. *Psychopharmacology* **236**, 2593–2611 (2019).
29. Scheggia, D., Bebensee, A., Weinberger, D. R. & Papaleo, F. The ultimate intra-/extra-dimensional attentional set-shifting task for mice. *Biol. Psychiatry* **75**, 660–670 (2014).
30. Scarsi, F., Scheggia, D. & Papaleo, F. Automated two-chamber operon ID/ED task for mice. *Curr. Protoc. Neurosci.* **94**, e109 (2020).
31. David, D. J. *et al.* Neurogenesis-dependent and -independent effects of fluoxetine in an animal model of anxiety/depression. *Neuron* **62**(4), 479–493 (2009).
32. Hurtubise, J. L. & Howland, J. G. Effects of stress on behavioral flexibility in rodents. *Neuroscience* **345**, 176–192 (2017).
33. Colacicco, G., Welzl, H., Lipp, H. P. & Würbel, H. Attentional set-shifting in mice: Modification of a rat paradigm, and evidence for strain-dependent variation. *Behav. Brain Res.* **132**, 95–102 (2002).
34. Proietti Onori, M., Ceci, C., Laviola, G. & Macri, S. A behavioural test battery to investigate tic-like symptoms, stereotypies, attentional capabilities, and spontaneous locomotion in different mouse strains. *Behav. Brain Res.* **267**, 95–105 (2014).
35. Oliver, Y. P., Ripley, T. L. & Stephens, D. N. Ethanol effects on impulsivity in two mouse strains: Similarities to diazepam and ketamine. *Psychopharmacology* **204**, 679–692 (2009).
36. Asinof, S. K. & Paine, T. A. The 5-choice serial reaction time task: A task of attention and impulse control for rodents. *J. Vis. Exp.* **90**, e51574 (2014).
37. Carli, M., Robbins, T. W., Evenden, J. L. & Everitt, B. J. Effects of lesions to ascending noradrenergic neurones on performance of a 5-choice serial reaction task in rats; implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal. *Behav. Brain Res.* **9**, 361–380 (1983).
38. Robbins, T. W. The 5-choice serial reaction time task: Behavioural pharmacology and functional neurochemistry. *Psychopharmacology* **163**, 362–380 (2002).
39. Humby, T., Wilkinson, L. & Dawson, G. Assaying aspects of attention and impulse control in mice using the 5-choice serial reaction time task. *Curr. Protoc. Neurosci.* **Chapter 8**, Unit 8 5H (2005).
40. Sultana, R., Ogundele, O. M. & Lee, C. C. Contrasting characteristic behaviours among common laboratory mouse strains. *R. Soc. Open Sci.* **6**, 190574 (2019).
41. Wallace, J., Marston, H. M., McQuade, R. & Gartside, S. E. Evidence that aetiological risk factors for psychiatric disorders cause distinct patterns of cognitive deficits. *Eur. Neuropsychopharmacol.* **24**, 879–889 (2014).
42. Nikiforuk, A. & Popik, P. Long-lasting cognitive deficit induced by stress is alleviated by acute administration of antidepressants. *Psychoneuroendocrinology* **36**, 28–39 (2011).
43. Ng, C. W., Noblejas, M. I., Rodefer, J. S., Smith, C. B. & Poremba, A. Double dissociation of attentional resources: Prefrontal versus cingulate cortices. *J. Neurosci.* **27**, 12123–12131 (2007).
44. Garner, J. P., Thogerson, C. M., Würbel, H., Murray, J. D. & Mench, J. A. Animal neuropsychology: Validation of the intra-dimensional extra-dimensional set shifting task for mice. *Behav. Brain Res.* **173**, 53–61 (2006).
45. Szlachta, M. *et al.* Effect of clozapine on ketamine-induced deficits in attentional set shift task in mice. *Psychopharmacology* **234**, 2103–2112 (2017).
46. Owen, A. M., Roberts, A. C., Hodges, J. R. & Robbins, T. W. Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. *Brain* **116**, 1159–1175 (1993).
47. Roberts, A. C., Robbins, T. W. & Everitt, B. J. The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *Q. J. Exp. Psychol. B* **40**, 321–341 (1988).
48. Trobalon, J. B., Miguez, D., McLaren, I. P. L. & Mackintosh, N. J. Intradimensional and extradimensional shifts in spatial learning. *J. Exp. Psychol. Anim. B* **29**, 143–152 (2003).
49. Cheng, J. T. & Li, J. S. Intra-orbitofrontal cortex injection of haloperidol removes the beneficial effect of methylphenidate on reversal learning of spontaneously hypertensive rats in an attentional set-shifting task. *Behav. Brain Res.* **239**, 148–154 (2013).
50. Scheggia, D. & Papaleo, F. An operant intra-/extra-dimensional set-shift task for mice. *J. Vis. Exp.* **107**, e53503 (2016).

## Acknowledgements

We wish to thank Flavia Chiarotti for statistical advice; Luigia Cancemi and Isabella Marcucci for valuable assistance with animal care. Ugo Basile S.r.l. (Gemonio, Italy) is gratefully acknowledged for supplying the apparatus (Operon 49500) on approval for Experiment 1 and for precious technical assistance.

## Author contributions

S.M., E.P. and F.Z. designed the research; E.P. and F.Z. supervised the experiments; C.S., C.B., A.M.O. and M.P. performed the experiments; F.Z. and E.P. performed the statistical analyses and prepared all figures; S.M., V.S., E.P. and F.Z. interpreted the results; E.P. and F.Z. wrote the first draft of the manuscript; S.M., V.S., E.P. and F.Z. revised and consolidated the manuscript. All authors reviewed and approved the final version of the manuscript.

## Funding

This work received financial support by the Italian Space Agency (ASI), project “In-situ Resource Bio-Utilization per il supporto alla vita nello Spazio” (ReBUS) to SM, by the Italian Ministry of Health, project GR-2016-02361921 to SM and VS. This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 847879.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-43631-8>.

**Correspondence** and requests for materials should be addressed to S.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023