

---

# MT-aIAT: Integrating mouse tracking into memory-detection aIAT

---

## *Authors:*

Xinyi XU

Centre for Cognitive and Brain Sciences and Department of Psychology,  
University of Macau

School of Automation, Wuhan University of Technology

Xianqing LIU

Department of Psychology, University of Macau

Xiaoqing HU

Department of Psychology, The University of Hong Kong

Haiyan WU\*

Centre for Cognitive and Brain Sciences and Department of Psychology,  
University of Macau

\* Corresponding author: Haiyan Wu

Centre for Cognitive and Brain Sciences and Department of Psychology,  
University of Macau, Taipa, Macau, China

Address: N21-G004d, University of Macau, Taipa, Macau

E-mail: haiyanwu@um.edu.mo (H. Wu)

\* Word count: 7677 words

# MT-aIAT: Integrating mouse tracking into memory-detection autobiographical Implicit Association Test

---

## Abstract

This study assesses the validity of a newly integrated memory detection method, MT-aIAT, which is a combination of the autobiographical Implicit Association Test (aIAT) and the mouse-tracking method. Participants were assigned to steal a credit card and then performed the aIAT while mouse tracker was recording their motor trajectories. Replicating previous work, we found a RT congruency effect. Critically, the mouse trajectories indicate a congruency effect and a block order effect, suggesting the validity of mouse-tracking technique in unraveling real-time measurement of the IAT congruency effect. Lastly, to test the computational modeling in MT-aIAT, we posited a connectionist model combined with the drift-diffusion model to simulate participants' behavioral performance. Our model captures the ubiquitous implicit bias towards the autobiographical event. Implications of the MT-aIAT in identifying autobiographical memories, the combination of MT-aIAT with computational modeling approach were discussed.

*Keywords:* concealed memory detection, autobiographical IAT, mouse-tracking, computational modeling

---

## 1. Introduction

Though a lie be well drest, it is ever overcome. The attempts to detect deception and concealed memory by physiological and behavioral measurements have evolved for more than a hundred years, captured a broad interest from fields including psychology, forensic, neuroscience and even ethics (Agosta & Sartori, 2013; Chassot et al., 2015; Verschuere et al., 2011; Vrij & Fisher, 2016; Nahari, 2018). In this study, we focused on the autobiographical Implicit Association Test (aIAT) (Sartori et al., 2008; Agosta et al., 2011), one of the recently developed behavioral procedures that bear promises in memory detection. Specifically, we proposed a modified computer-based aIAT paradigm, to be combined with mouse tracking, a popular response measurement technique as a mouse-tracking aIAT (hereinafter MT-aIAT). We applied the MT-aIAT in a mock crime scenario to demonstrate how the aIAT can detect concealed autobiographical memories, and we employed computational models to reveal cognitive mechanisms underlying aIAT memory detection.

The aIAT is an adaptation of the Implicit Association Test (IAT) (Greenwald et al., 1998). Unlike IAT that has been widely used in social cognition research examining attitudes, the aIAT aims to identify the veracity of autobiographical memories by assessing the associative strength between subjects' autobiographical memories and objective events. Specifically, the aIAT requires participants to categorize statements belonging to one of two categories – objectively true or false statements (e.g., “I am in the experiment room” vs. “I am in a shop”) and autobiographical statements that only half is true based on participants' experience (e.g., “I stole a credit card” vs. “I copied the confidential information from the computer”). The aIAT effect is reflected by differences in reaction time (RT), and is usually quantified by D scores defined as the differences of the averaged RTs between incongruent and congruent conditions, divided by the pooled standard deviation  $((RT_{congruent} - RT_{incongruent})/SD)$  (Gonick et al., 1993). Larger D scores would represent stronger aIAT effects (Greenwald et al., 1998), such that participants would associate one type of autobiographical statements with

truth. Notably, the order of congruent and incongruent blocks may influence the IAT effect, such that the IAT effect is larger when the congruent block is performed first (Greenwald et al., 1998). When applying IAT method in empirical studies, this order effect should be considered.

Though the RT-based aIAT has been validated in memory detection, it is not without limitations. For example, several studies show that manipulating instruction of training or response strategies can beat the traditional aIAT (Hu et al., 2012; Verschuere et al., 2009). Also, RT-based aIAT does not reveal the real-time categorization processes, given that only the final behavioral output (RT, accuracy) is recorded (Yu et al., 2012; Smeding et al., 2016). A promising approach to complete the aIAT is to provide a real-time measurement of categorization by adding mouse-tracking indices (Duran et al., 2010). Analysis of mouse-tracking data can capture the temporal dynamics of memory categorization in terms of spatiotemporal patterns. As the mouse-tracking method assumes that every individual has distinctive spatial and temporal motor patterns and habits of computer mouse usage (Sartori et al., 2018), it provides individuals' real-time categorization dynamics along a millisecond temporal scale that allow the investigation of cognitive processing underlying memory detection (Freeman et al., 2008, 2011). Meanwhile, strategic responding can be tracked in the mouse trajectories, as the initial attraction to the other response should be observable in the response curve (Sartori et al., 2018).

Recent work has successfully incorporated mouse-tracking with IAT. For example, Yu et al. (2012) integrated the mouse tracking technique to a flower-insect IAT and two implicit self-esteem IATs, which showed both classical RT-based IAT effects and the potential of mouse trajectories in revealing the underlying process of IAT. Monaro et al. (2021b) also demonstrated the effectiveness of mouse-tracking IAT to assess implicit preferences towards social networks such as Facebook and Twitter, extending the MT-IAT to a novel field such as consumer research. The covert nature of mouse tracking and the real-time, continuous motor trajectories are also instrumental in memory detection tasks (Sartori et al., 2018; Papesh & Goldinger, 2012). For instance,

in a word recognition task, participants made new/old decisions while being tracked to their mouse coordinates and then underwent a confidence assessment (Papesh & Goldinger, 2012). By examining response trajectories and subsequent confidence, the researchers found stronger memories corresponded to fast linear movements, while weaker memories produced slower curvilinear movements. Taken together, mouse-tracking could provide continuous spatiotemporal information in assessing attitudes and memory strength. Based on mouse tracking’s advantage in unraveling real-time mouse trajectories, we combined mouse-tracking with aIAT to test the effectiveness of the mouse tracking-based aIAT method in detecting autobiographical memories.

Studies have also found discrepant stimuli modality effect on IAT effect size (e.g., Dasgupta et al., 2000; Foroni & Bel-Bahar, 2010). Picture-IATs, in general, are associated with a smaller IAT effect than word-IATs (Nosek et al., 2002; Mitchell et al., 2003). Although previous studies asked participants to lie to words (Wu et al., 2009) or pictures (Dong et al., 2010), few studies compared different modalities in memory detection tests. In our MT-aIAT, we will use both picture and word stimuli to examine the detection efficiency for different modalities (picture vs. word). The finding would contribute to the refinement of MT-aIAT method as a prerequisite for practical application.

RTs and mouse-tracking data of MT-aIAT can also be incorporated with computational modeling. A commonly used model in RT-based choices is the drift-diffusion model (DDM), which has been proved to be a powerful method for revealing internal properties of decision making process of both human and rodents (Brunton et al., 2013). In traditional DDM, the information begins at the postulated position  $z$  and accumulates with time at a speed  $v$ . The accumulation of information includes systematic as well as random influences. Decision is made when the accumulated evidence reaches the threshold (Klauer et al., 2007). The DDM parameters are gained through data fitting and have been implicated in explaining RTs during value-based decision making under low and high time pressure (Pedersen et al., 2017). To our knowledge, the IAT has only been modeled with the traditional DDM (Klauer et al., 2007; Van Ravenzwaaij

---

et al., 2011), which is not able to connect with mouse tracking trajectories, because it mainly focuses on simulating the distribution of RTs and the fitted parameters do not have temporal dynamics. Following Wong’s (2007) work, we focuses on temporal neural dynamics of decision processes and accurately simulated the IAT effect, allowing us to combine it with the mouse trajectories and disassociate the sensory information and motor responses. Further, the connectionist model quantifies the sensory information which bridge the gap between the non- value-based experiment (MT-aIAT) with this precisely described model, while the DDM models the behavioral performances. Therefore, we would like to present the application of computational modeling in the MT-aIAT method, and offer us a top-down computational basis of implicit bias and explicit RT difference

In summary, the specific objective of the current study is to: 1) validate the MT-aIAT method by tracking both the classic IAT effect index and the mouse trajectories, and 2) test potential computational modeling approach to shed light on the underlying mechanism of MT-aIAT decision process. We assume that MT-aIAT is able to perform well on autobiographical memory detection based on either RTs or mouse dynamic indices. It is predicted that the IAT block order effect will still exist in MT-aIAT, and that word stimuli will evoke a stronger IAT effect than picture stimuli as previous findings showed. Our work goes significantly beyond the current literature by providing a newly adopted memory detection paradigm integrating mouse-tracking with the aIAT, with a more comprehensive analysis of the IAT effect and relevant mouse-tracking dynamics considering the effect of block order and stimulus modality in aIAT. Besides, we will attempt to use computational modeling for MT-aIAT simulation and expect that our connectionist model can simulate the implicit bias consistent with our empirical data by integrated with DDM.

## 2. Methods

### 2.1. Participants

59 healthy undergraduate and graduate students (27 males,  $M_{age} = 21.36$ ,  $SD = 2.32$ ) participated in this experiment. They were divided into congruent block first MT-aIAT group ( $n = 33$ , 16 males,  $M_{age} = 21.48$ ,  $SD = 2.67$ ) and incongruent block first MT-aIAT group ( $n = 26$ , 11 males,  $M_{age} = 21.19$ ,  $SD = 1.84$ ) to balance and investigate the effect of block order. To achieve greater than 80% power to detect a large effect of  $d = 0.80$  at  $\alpha = 0.05$  in our analyses of paired t-tests, we calculated the minimal sample size, which was 14, using the package `pwr` in R (Champely et al., 2018). We oversampled and recruited 33 participants for the congruent block first group and 26 participants for the incongruent block first group. All the participants were right-handed with normal or corrected-to-normal vision, and had not participated in any similar studies. Each participant signed informed consent prior to the formal experiment, and the experimental protocol was approved by the local ethical review committee. At the end of the study, the participants were paid 50-60 CNY.

### 2.2. Procedure

#### 2.2.1. Mock Crime Session

The mock crime setting has been used in existed aIAT studies (e.g., Sartori et al., 2008; Verschuere et al., 2009; Agosta et al., 2011; Hu et al., 2012). Participants were asked to select one out of two envelopes deciding which task they were going to perform. They were informed that one of the tasks was to steal a credit card in a wallet, the other was to copy a confidential file on a computer. The content in the two envelopes was the same – to steal the credit card (see Verschuere et al., 2009), by which to bring participants a sense of involvement to ‘commit the crime’ and drive them to perform the task with self-motivation. After revealing the task in envelope, they were guided to another lab room (mock crime room) to find the wallet and steal the credit card. To increase

the ecological validity, the experimenter informed participants before the session that they had to withdraw from the experiment if caught during the mock crime session.

### 2.2.2. MT-aIAT Task

After the mock crime session, participants were arranged to sit in front of a monitor in another testing room to complete the aIAT session. The aIAT was performed using a procedure analogous similar to previous work (Sartori et al., 2008; Marini et al., 2016), while implemented with a mouse tracker to record the mouse trajectories. The stimuli were presented through Mousetracker (<http://www.mousetracker.org/>), which recorded the mouse position about 70 times per second (70 Hz). Participants were instructed to click START at the bottom center of the screen for each trial, then the event stimulus showed up in the center. They were instructed to classify the stimulus by clicking the key at the left upper corner of the screen (R1) or the key at the right upper corner (R2). Participants were required to move the mouse quickly and accurately in the task; otherwise, a reminder would appear to urge them to respond as quickly as possible. The aIAT was structured in seven blocks, including three simple blocks (block 1, 2, and 5) and four combined blocks (block 3, 4, 6, and 7):

- Block 1 (20 trials) required all participants to make a binary classification based on the stimuli’s logical attributes: they were asked to discriminate whether the sentence displayed was logically true (e.g., “I am on the third floor”) or false (e.g., “I am in a shop”), and click the corresponding key (R1 for true and R2 for false).
- In block 2 (20 trials) and block 5 (20 trials), participants classify the stimulus only depending on whether it was associated with the crime-relevant event (event 1) they had committed (e.g., “I steal a credit card”) or the crime-irrelevant event (event 2) (e.g., “I copied the confidential information from the computer”), and click the corresponding corner/area. For the congruent block first MT-aIAT group, R1 corresponded to event



1(the relevant criminal event), and R2 corresponded to event 2 in block2. In block 5, the response pattern was reversed: R1 corresponded to event 2, and R2 corresponded to event 1. For the incongruent block first MT-aIAT group, R1 corresponded to event 2, and R2 corresponded to event 1 in block 2; while in block 5, the response pattern was reversed.

- Block 3 (60 trials), block 4 (200trials), block 6 (60 trials), and block 7 (200 trials) were the combined blocks, requiring participants to categorize both “objectively true or false” events and “crime -relevant or -irrelevant” events. The four combined blocks were subdivided into congruent blocks and incongruent blocks. In the congruent blocks, event 1 and the logically true sentences shared the same motor response R1, while stimuli related to event 2 and logically false sentences shared the same motor response R2. In the incongruent blocks, the respondents learn a reversal of response assignment where the two combinations change to ‘crime-relevant + false’ and ‘crime-irrelevant + true’. For the congruent block first group, block 3 and block 4 were congruent, and block 6 and block 7 were incongruent (see Figure 2(a)). For the incongruent block first group, block 3 and block 4 were incongruent, and block 6 and block 7 were congruent (see Figure 3(a)).

With the aim to examine the stimulus type effect in MT-aIAT, we implemented both the picture and word stimuli for events. The stimuli set of picture MT-aIAT included pictures related to the mocked criminal event and the stimuli set of word MT-aIAT included short word descriptions of the mocked criminal event (e.g., open the door). A complete set of stimuli used in the task are presented in Supplementary Materials 1. All the participants were exposed to both types of stimuli in the aIAT task at random.

The MT-aIAT task lasted for about 20 minutes. During the task, x and y coordinates along the cursor trajectories were recorded.

### 2.3. Statistical Analysis

Following Greenwald et al.(1998), we analyzed data obtained from block 4 and block 7. Participants with an error rate higher than 30% were excluded from further analysis (one of the subjects was excluded for the high error rate of 44%). To reduce the noise caused by trials where participants did not comply with the instructions, we removed the trials with the error or slow response (RTs > 5000ms, see [Anikin & Johansson, 2019](#)).

All analysis of behavioral data was done through R (version 4.1). Each participant’s D score of aIAT was calculated according to the participant’s behavioral performance differences between the congruent blocks and incongruent blocks. To investigate the IAT effect, we firstly ran paired t-tests on RTs and mouse geometric indices between congruent and incongruent trials of the two groups separately. Then we performed a three-way repeated-measures analysis of variance (ANOVA) on RTs to explore potential effects, with conditions (congruent vs. incongruent), event types (the crime-relevant event vs. the crime-irrelevant event), and stimulus modalities (word vs. picture) as independent variables. A two-way ANOVA with event type and logical attributes (true/false) as independent variables on RTs was performed to particularly examine the impact of event type and congruency that we focus on the most.

### 2.4. Simulations of Reaction Time

#### 2.4.1. Connectionist Model

The connectionist model is comprised of four neural sub-populations that respond to the perception of four features, including true (True), false (False), crime-relevant (CR), and crime-irrelevant (CI) within the agent (see Figure 4(a)) ([Bedder et al., 2019](#)). When the agent perceives its own features, a neuromodulatory signal allows synaptic connections between active sub-populations to be strengthened by a Hebbian learning rule (see Figure 4(a)) ([Hebb, 2005](#)). In this case, the simulated agent committed crime-relevant events as our participants did, such that the connectionist model comes to encode strong associations between neurons encoding true and crime-relevant features. During subsequent

perception, sensory input to sub-populations encoding the features of that agent generates additional activity in the network via recurrent synaptic connections if those features overlap with the encoded features (see Figure 4(c, d)). For detailed algorithms, see Supplementary Materials 2.

#### 2.4.2. Drift-Diffusion Models (DDM)

Behavioral performance on the IAT can be modeled with DDM (Wong et al., 2007; Klauer et al., 2007; Van Ravenzwaaij et al., 2011) consisting of two self-excitatory but mutually inhibitory neural populations (Bedder et al., 2019). These two neural populations code for left and right motor outputs respectively (see Figure 5(a)). External sensory evidence integrated with noise accumulates until the firing rate of one population reaches a pre-defined decision threshold (Bogacz et al., 2006; Koop & Johnson, 2013; Ratcliff & Rouder, 2000). The time taken to reach the decision threshold produces an RT, while the winning population corresponds to the decision made. In our simulations, the sensory evidence provided to each DDM motor population in each IAT trial is determined by activity in the connectionist model (see Figure 5). Neurons coding for the IAT stimulus receive a set level of external sensory input, while additional input to either motor response population arises from recurrent excitation within the connectionist model. Obviously, the level of the total input of the left neural population in the congruent condition is larger than the right one or either in the incongruent condition (indicated by thicker colored arrows in Figure 5). As a result, the left neural population in the congruent condition is the earliest one to reach the threshold; thus, the participants were most likely to choose the left option, and the RTs of congruent condition was significantly smaller than that of incongruent condition. For detailed algorithms, see Supplementary Materials 2.

## 2.5. Mouse-Tracking Data Analysis

### 2.5.1. Mouse Trajectory Preprocessing

Analogous with statistical analysis, we extracted data from block 4 and block 7, and excluded subject 106 for his high error rate (44%) and removed the trials with the error or slow response (RTs > 5000ms) (Anikin & Johansson, 2019). Standard mouse-tracking preprocessing was conducted temporally and spatially (Freeman & Ambady, 2010). We rescaled all responding mouse trajectories into standard coordinate space (top left: [-1, 1]; top right: [1,1]) so that the cursor always started at [0,0]. The duration of each trial was sliced into 101 identical time bins using linear interpolation to permit the average of their length across multiple trials (see Spivey et al., 2005; Dale et al., 2007; Freeman et al., 2008, 2010; Duran et al., 2010; Sullivan et al., 2015).

### 2.5.2. Mouse Trajectory Measurements

To get a trial-by-trial category co-activation index, we calculated the maximum perpendicular deviation (MD), the area under the curve (AUC) and the mean maximum deviation (MAD) of each mouse trajectory by R package Mouse-trap (Kieslich & Henninger, 2017). In the two-choice mouse-tracking tasks, deviation in the mouse trajectory of a subject toward an opposite category response (indexed by MD) is a well-validated measure of the extent to which that other category is also activated during the perceptual decision process (Spivey & Dale, 2006; Freeman & Ambady, 2010; Freeman et al., 2011).

### 2.5.3. Temporal Analysis

We averaged all mouse trajectories into four conditions (congruent and the crime-relevant event, congruent and the crime-irrelevant event, incongruent and the crime-relevant event, incongruent and the crime-irrelevant event) after preprocessing the mouse trajectories. To test the significant temporal difference between trajectories statistically, we calculated the positions along with time bins of every trial in two conditions (congruent vs. incongruent) and compared them through paired t-test (Chemin et al., 2018; Kieslich & Henninger, 2017).

Besides, we split trials into crime-relevant and crime-irrelevant events, further examined the discrepancy between the two conditions (congruent vs. incongruent).

#### 2.5.4. Trajectory Analysis

A linear regression is performed to examine how the trajectory angle depends on each stimulus of different conditions in every normalized time point (Sullivan et al., 2015). We normalized the trajectories angles so that  $+45^\circ$  always indicated a direct movement to the right,  $0^\circ$  indicated a movement straight upward, and  $-45^\circ$  indicated a direct movement to the left. To quantify the stimulus intensity, we made use of the outputs of the connectionist model where stimuli related to the crime-relevant or true events induce higher firing rates. These regressions were estimated at the individual level. The relevant estimated coefficients ( $\beta$ ) were pooled across congruent and incongruent conditions.

### 3. Results

#### 3.1. Results from Merged Two Groups

The data for the following analysis was based on RTs and mouse indices in block 4 and block 7 in both congruent and incongruent block first groups. The error rate of the congruent condition was 0.044 and the incongruent condition was 0.25.

##### 3.1.1. RTs and D Score

To analyze the RTs of the current MT-aIAT, we conducted analysis including an ANOVA and paired t-tests. For the ANOVA, we firstly considered the block order (congruent or incongruent block first) as between-subjects variable, condition (congruent or incongruent) and event type (crime-relevant or crime irrelevant event) as within-subjects variables. The results indicated that no significant main effect of block order on RTs was found ( $F(1,57) = 0.58$ ,  $p = 0.45$ ,  $\eta_p^2 = 0.010$ ), but a significant main effect of condition ( $F(1,57) = 92.81$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.156$ ). There was also an interaction between the order and the

condition ( $F(1, 57) = 30.03, p < 0.001, \eta_p^2 = 0.056$ ) (see Figure 6). The post hoc analysis showed that both orders indicated a strong IAT effect (i.e., longer RTs for incongruent block. Congruent block first:  $t = -3.17, p = 0.0027$ ; incongruent block first:  $t = -10.00, p < 0.001$ ), and incongruent block first group showed a bigger IAT effect ( $t = 5.48, p < 0.001$ ). Also, longer RTs in incongruent condition than congruent condition were found as expected ( $p < 0.001$ ), for both the crime-relevant and crime-irrelevant events (see Figure 1(b)). Furthermore, RTs to picture stimuli were generally shorter than to word stimuli ( $t(57) = -9.93, p < 0.001, 95\% CI = -218.24$  to  $-144.97, Cohen's d = 1.30$ ) (picture-congruent:  $M = 1195.27, SD = 218.23$ ; picture-incongruent:  $M = 1433.89, SD = 416.15$ ; word-congruent:  $M = 1359.05, SD = 276.71$ ; word-incongruent:  $M = 1592.95, SD = 362.10$ ) (see Figure 1(a)).

Weighted D scores of each participant were calculated across block 4 and block 7 (for a detailed algorithm, see [Greenwald et al., 2003](#)). A negative D score means that RTs were faster in congruent condition than incongruent condition, and a bigger absolute value of D score represents a stronger IAT effect. Since the connectionist model generated different firing rates of task-relevant items and task-irrelevant items in different conditions ([Bedder et al., 2019](#)), sensory evidence, which was proportional to the difference in firing rates of different conditions, would be diverse, thus affected the accumulation process in DDM and the final RTs. We investigated the IAT effect of crime-relevant and crime-irrelevant, respectively. Specifically, there was no significant discrepancies of D score between crime-relevant and crime-irrelevant ( $t(57) = -0.95, p = 0.35, 95\% CI = -0.15$  to  $0.053, Cohen's d = 0.12$ ). But crime-relevant event possessed a stronger IAT effect (D score of the crime-relevant event:  $M = -0.37, SD = 0.36$ ; D score of the crime-irrelevant event:  $M = -0.32, SD = 0.45$ ). A significant IAT effect of both word and picture stimuli was also observed (D score of word stimuli:  $M = -0.34, SD = 0.39, t(57) = -6.48, p < 0.001, 95\% CI = -0.44$  to  $-0.23, Cohen's d = 1.01$ ; D score of picture stimuli:  $M = -0.38, SD = 0.41, t(57) = -7.09, p < 0.001, 95\% CI = -0.49$  to  $-0.27, Cohen's d = 1.01$ ), but there was no significant difference on IAT effect between the two modalities ( $t(57) =$

1.00,  $p = 0.32$ , 95%  $CI = -0.14$  to  $0.046$ , *Cohen's d* = 0.13).

### 3.1.2. Mouse Indices

We compared the MAD values of two conditions (congruent:  $M = 0.44$ ,  $SD = 0.20$ ; incongruent:  $M = 0.51$ ,  $SD = 0.19$ ) and observed a significant congruency effect on MAD ( $t(57) = -3.45$ ,  $p < 0.001$ , 95%  $CI = -0.11$  to  $-0.028$ , *Cohen's d* = 0.45). Other results considering mouse indices are illustrated in the Figure 1(c-e). Similar to RTs analysis, the analysis of mouse indices also separately examines the IAT effect in the processing of crime-relevant and crime-irrelevant event. Significant differences of MAD and AD both in the crime-relevant event between congruent and incongruent blocks were shown by the results (MAD:  $t(57) = -2.57$ ,  $p = 0.013$ , 95%  $CI = -0.15$  to  $-0.018$ , *Cohen's d* = 0.34; AD:  $t(57) = -2.62$ ,  $p = 0.01$ , 95%  $CI = -0.047$  to  $-0.0062$ , *Cohen's d* = 0.37), while there was no significant discrepancy for AUC in any pair of comparisons.

### 3.1.3. Trajectory Analysis Results

In the temporal trajectory analysis, the aggregated trajectories of two blocks (see Figure 7(c)) and two events in each block (see Figure 7(f)) were derived. We performed paired t-tests on each normalized time point of two conditions using R package Mousetrap (Kieslich & Henninger, 2017). The t-tests showed that trajectories of two conditions had a significant discrepancy from the 50th to 93rd time point (see Figure 8(c)). We then focused on the trajectories of each event (see Figure 8(f)) and found a significant discrepancy only when processing the crime-irrelevant event (the crime-irrelevant event: 17th - 86th time point).

We also did a regression analysis for the sensory evidence. The estimated slope ( $\beta$ ) was illustrated in Figure 9(c) across the 101 time windows. The result showed that there was no significant difference between the  $\beta$ s under two conditions .

#### 3.1.4. Simulated Reaction Time

Our study simulated 59 agents with a connectionist model which consists of  $N = 4$  neurons that code for the crime-relevant, the crime-irrelevant, true, and false events. We assumed that each neuron only encodes a single feature rather than a conjunction of multiple features, because of simplicity only – the conjunctive coding of features in the connectionist network would have no qualitative effect on the results. Finally, we did observe the IAT effect in the simulation and calculated the IAT score (IATscore = -0.22), which was close to the IAT score of the congruent block first group (IATscore = -0.19). Compatible with the empirical data, we demonstrated that simulated agents exhibit significantly negative IAT scores, suggesting smaller RTs in congruent trials. Finally, we regressed the sensory evidence on the trajectory angle in each time point to investigate the impact of the stimuli on motor dynamics (Sullivan et al., 2015). As shown in Figure 9, in the congruent block first group, stimuli in congruent trials evoked a stronger effect on mouse trajectories than in incongruent trials at all time steps; in the incongruent block first group, stimuli in incongruent trials had a stronger impact in the very early time step, but later it was exceeded by congruent trials.

#### 3.1.5. The Block Order Effect

According to Greenwald et al. (1998), the IAT effect in the congruent block first and incongruent block first condition may differ (Greenwald et al., 1998). Inconsistent with previous studies, our results indicated a larger IAT effect for incongruent block first condition (congruent block first: IATscore = -0.19; incongruent block first: IATscore = -0.49). To demonstrate the impact of block order on the aIAT effect, we further reported the independent analysis on the congruent block first group and incongruent block first group.

### 3.2. Congruent Block First Group

We observed an overall high accuracy of MT-aIAT in two congruency conditions (error rate = 0.12) and higher error rates of in incongruent condition than



congruent condition (congruent block: 0.03; incongruent block: 0.22).

### 3.2.1. Reaction Time and D Score

Paired t-tests were performed to compare the RTs in the congruent and incongruent conditions. We found significant shorter RTs for congruent association condition than incongruent association condition as expected ( $t(31) = -2.13$ ,  $p = 0.041$ , 95%  $CI = -179.326272$  to  $-3.944326$ , *Cohen's d* = 0.38), for both the crime-relevant and the crime-irrelevant events stimuli (see Figure 2(c)). Furthermore, RTs to picture stimuli were generally smaller than to word stimuli ( $t(31) = -6.78$ ,  $p < 0.001$ , 95%  $CI = -236.93$  to  $-127.29$ , *Cohen's d* = 1.20) (see Figure 2(b)).

Weighted D scores (Greenwald et al., 2003) of each participant were calculated across block 4 and block 7. Specifically, there was no significant discrepancies of D score between crime-relevant and crime-irrelevant condition ( $t(31) = -1.98$ ,  $p = 0.057$ , 95%  $CI = -0.28$  to  $0.0041$ , *Cohen's d* = 0.35), but crime-relevant event possessed a stronger IAT effect (D score of the crime-relevant event:  $M = -0.31$ ,  $SD = 0.40$ ; D score of the crime-irrelevant event:  $M = -0.18$ ,  $SD = 0.45$ ). Also, we observed a significant IAT effect on word and picture stimuli (D score of word stimuli:  $M = -0.28$ ,  $SD = 0.42$ ,  $t(31) = -3.74$ ,  $p < 0.001$ , 95%  $CI = -0.4384126$  to  $-0.1289101$ , *Cohen's d* = 0.66; D score of picture stimuli:  $M = -0.22$ ,  $SD = 0.38$ ,  $t(31) = -3.12$ ,  $p < 0.001$ , 95%  $CI = -0.35879628$  to  $-0.07531092$ , *Cohen's d* = 0.55), but there was no significant difference between them ( $t(31) = -1.12$ ,  $p = 0.27$ , 95%  $CI = -0.055$  to  $0.19$ , *Cohen's d* = 0.20). Overall, these results showed that participants tended to respond more quickly to picture stimulus in congruent blocks (picture congruent:  $M = 1240.68$ ,  $SD = 187.48$ ; picture incongruent:  $M = 1300.93$ ,  $SD = 254.24$ ; word congruent:  $M = 1396.69$ ,  $SD = 250.08$ ; word incongruent:  $M = 1532.55$ ,  $SD = 340.63$ ).

### 3.2.2. Mouse Indices

To analyze mouse indices, we calculated the AD values in two conditions (congruent:  $M = 0.12$ ,  $SD = 0.055$ ; incongruent:  $M = 0.13$ ,  $SD = 0.058$ ) and performed a paired t-test comparing them where we observed a significant discrepancy of AD in two conditions ( $t(31) = -2.34$ ,  $p = 0.026$ , 95%  $CI = -0.029$  to  $-0.0020$ , *Cohen's d* = 0.41). Paired t-tests were also conducted comparing AUC, MAD and AD, respectively, between different event type. The results were illustrated in Figure 2(d-e) below. There was no significant AUC discrepancy found. However, for MAD and AD, their values when responding to event 1 were significantly larger than event 2 in incongruent blocks (MAD:  $t(32) = 2.24$ ,  $p = 0.033$ , 95%  $CI = 0.0093$  to  $0.21$ , *Cohen's d* = 0.42; AD:  $t(32) = 2.36$ ,  $p = 0.025$ , 95%  $CI = 0.0045$  to  $0.063$ , *Cohen's d* = 0.44). These results together suggested MAD and AD might be good predictors of the autobiographical event.

### 3.2.3. Trajectory Analysis Results

In the temporal trajectory analysis, the aggregated trajectories of two blocks (see Figure 7(a)) and two events in each block (see Figure 7(d)) were derived. The paired t-tests were performed comparing two conditions on each normalized time point using R package Mousetrap (Kieslich & Henninger, 2017). The results showed that trajectories of two conditions had a significant discrepancy from the 53th to 94th time point (see Figure 8(a)). We then focused on the trajectories of each event (see Figure 8(d)) and found that when accepting the irrelevant event (the crime-irrelevant event) as the stimulus, the number of time steps on which the trajectories of two conditions had a significant difference was more than the crime-relevant event (the crime-relevant event: 56th-93rd time point; the crime-irrelevant event: 14th-44th, 67th-84th time point). For each normalized time window and each individual, the local angle trajectory was regressed against the sensory evidence conducted by the connectionist model. The absolute values of the estimated slope ( $\beta$ ), illustrated in Figure 9(a) across the 101 time windows, provided a measure of the degree to which the value information affected the process of choice at different times. A larger absolute value of ( $\beta$ ) indicated a

stronger effect that the stimulus has on trajectory angle. The results showed that the absolute value of regression coefficient ( $\beta$ ) in congruent trials was larger in all time points, indicating that stimuli under this situation motivated stronger resonance during the process of choice before the final decision.

### 3.3. Incongruent Block First Group

The following analysis was based on RTs and mouse indices in block 4 and block 7 in the incongruent block first group. The error rate of the congruent condition was 0.060, and incongruent condition was 0.29.

#### 3.3.1. Reaction Time and D Score

Paired t-tests were performed to compare the RTs in the congruent and incongruent conditions. RTs in the congruent association condition was found significantly shorter than in the incongruent association condition as expected ( $t(25) = -9.69$ ,  $p < 0.001$ , 95%  $CI = -340.19$  to  $-220.92$ , *Cohen's d* = 1.90), for both the crime-relevant stimuli or crime-irrelevant stimuli (see Figure 3(c). Consistent with congruent block first group, RTs to picture stimuli in incongruent block first group were generally smaller than word stimuli ( $t(25) = -7.39$ ,  $p < 0.001$ , 95%  $CI = -231.43$  to  $-130.54$ , *Cohen's d* = 1.45) (picture congruent:  $M = 1139.37$ ,  $SD = 243.11$ ; picture incongruent:  $M = 1613.09$ ,  $SD = 520.18$ ; word congruent:  $M = 1312.74$ ,  $SD = 304.93$ ; word incongruent:  $M = 1622.67$ ,  $SD = 387.83$ ) (see Figure 3(b).

Weighted D scores (for a detailed algorithm, see Greenwald et al., 2003) of each participant were calculated across block 4 and block 7. Specifically, there was no significant discrepancies of D score between crime-relevant and crime-irrelevant ( $t(25) = 0.89$ ,  $p = 0.38$ , 95%  $CI = -0.080$  to  $0.21$ , *Cohen's d* = 0.18), but opposite to congruent-block first group, crime-irrelevant event possessed a stronger IAT effect (D score of the crime-relevant event:  $M = -0.44$ ,  $SD = 0.30$ ; D score of the crime-irrelevant event:  $M = -0.50$ ,  $SD = 0.38$ ). Also, we both observed significant IAT effects of word and picture stimuli (D score of word stimuli:  $M = -0.40$ ,  $SD = 0.34$ ,  $t(25) = -5.92$ ,  $p < 0.001$ , 95%  $CI = -0.54$  to

-0.26, *Cohen's d* = 1.16; D score of picture stimuli:  $M = -0.58$ ,  $SD = 0.33$ ,  $t(25) = -8.81$ ,  $p < 0.001$ , 95%  $CI = -0.72$  to  $-0.45$ , *Cohen's d* = 1.72), and there was also significant difference between them ( $t(25) = 3.00$ ,  $p = 0.0060$ , 95%  $CI = 0.06$  to  $0.31$ , *Cohen's d* = 0.59).

### 3.3.2. Mouse Indices

We calculated the MAD values of two conditions (congruent:  $M = 0.46$ ,  $SD = 0.22$ ; incongruent:  $M = 0.56$ ,  $SD = 0.20$ ) and performed a paired t-test where we observed a significant discrepancy of MAD in two conditions ( $t(25) = -2.86$ ,  $p = 0.0084$ , 95%  $CI = -0.17$  to  $-0.03$ , *Cohen's d* = 0.56). The results of other mouse indices were illustrated in Figure 3(d-e). There was only a significant discrepancy for AUC in the condition comparison ( $t(25) = -2.21$ ,  $p = 0.037$ , 95%  $CI = -0.08$  to  $-0.003$ , *Cohen's d* = 0.43). The MAD and AD values of the crime-irrelevant stimuli in incongruent blocks were significantly larger than crime-relevant event (MAD:  $t(25) = 3.49$ ,  $p = 0.0023$ , 95%  $CI = 0.088$  to  $0.35$ , *Cohen's d* = 0.76; AD:  $t(25) = 3.12$ ,  $p = 0.0054$ , 95%  $CI = 0.026$  to  $0.13$ , *Cohen's d* = 0.68).

### 3.3.3. Trajectory Analysis Results

In temporal trajectory analysis, the aggregated trajectories of two blocks (see Figure 7(b)) and two events in each block (see Figure 5(e)) were derived. The paired t-tests were performed comparing two conditions on each normalized time point using R package Mousetrap (Kieslich & Henninger, 2017). The results showed that trajectories of two conditions have a significant discrepancy from the 51st to 60th and 75th to 90th time point (see Figure 8(b)). We then focused on the trajectories of each event (see Figure 8(e)). It was found that there were more time steps with significantly different trajectories in the two conditions when stimulated by crime-unrelated events than when stimulated by crime-related events.

We also did a regression analysis for the sensory evidence. The estimated slope ( $\beta$ ) illustrated in Figure 9(b) across the 101 time windows. Opposite

to the congruent block first group, the absolute value of regression coefficient ( $\beta$ ) of incongruent trials had a peak and exceeded the congruent trials in the early time steps and then, decreased to a very small value. Interestingly, the estimated slope ( $\beta$ ) of the latter congruent trials had a peak at around the 60th time step, which might be due to the association reactivation between the autobiographical event and its attribute.

In summary, for the IAT effect in the two groups with a different order, incongruent block-first group indicated a stronger IAT effect than RTs and all mouse indices showed the congruency effects.

## 4. Discussion

### 4.1. Validation: Behavioral Results and Mouse Indices in MT-aIAT

In general, the present results confirmed the validity of the MT-aIAT. Our MT-aIAT approach replicated the classic RT-based congruency effects observed in RT results. Moreover, the analysis of the mouse dynamic indices also demonstrated the congruency effects in MT-aIAT. The overall IAT effect in RTs and the mouse geometric indices (e.g., MAD and AD) was well observed in the merged data, regardless of the order of in/congruent blocks and stimuli modality (see Table 1, Figure 2 - 3), indicating the robustness of MT-aIAT task in detecting autobiographical memories. We did not compare RT-aIAT and MT-aIAT directly as the MT-aIAT including multiple indices including RT, MAD, AUC and AD etc.

Furthermore, our approach established the feasibility of the mouse-tracking method, which can be conveniently integrated into an aIAT paradigm (e.g., Yu et al., 2012; Monaro et al., 2021b). In spite of providing a continuous and dynamic view of the cognitive process, there are several other advantages to use mouse dynamics as an indicator of memory detection. First, mouse-tracking can be administered covertly, without participants' knowledge of the memory detection purpose (Sartori et al., 2018). Second, studies have shown high accuracy of the false identity detection task based on mouse-tracking, indicating

the reliability of mouse-tracking-based memory detection in specific contexts (Monaro et al., 2017a,b). Third, the mouse-tracking technique can be easily used among large samples (Sartori et al., 2018). Based on the above advantages, mouse-tracking technology has some possible application prospects such as false identity detection, concealed criminal memories detection, and malingering in the clinical field (Rosenfeld, 2018; Monaro et al., 2021a). Further investigation is needed to test the effectiveness and accuracy of mouse-tracking in various possible application contexts and combination with different memory detection tasks.

#### 4.2. Important Factors: Block Order and Modality

Our MT-aIAT method uncovered a congruency effect after added two factors, block order and modality, even though these two factors had some impact on the IAT effect size. In the congruent block first group, there was a weaker aIAT effect under four kinds of comparisons than in the incongruent block first group. As other factors of the experiments have been well controlled, we attribute the difference to the block order effect of MT-aIAT. This finding is contrary to Greenwald et al. (1998), who found a stronger IAT effect in the congruent block first group. Such discrepancy could be attributed to the difference between the IAT and aIAT task.

The IAT is essentially a categorization task that probes participants' implicit attitudes through the process of categorizing different categories of stimuli, whereas detecting memory for autobiographical events with aIAT involves the memory retrieval rather than simple categorization. It also seems possible that this inconsistency is due to the impact of some procedural factors such as the combination of word and picture stimuli, and the number of practice trials. The relatively high error rate suggests that participants' proficiency on the task may have influenced the congruency effect. It seems that to increase the number of practice trials to reduce the block order effect. Also, this effect could be countered by giving pairings to different participants first when studying groups of people. The present methods design did not try to reduce the block order

effect but to find which mouse-tracking indicators still showed a significant IAT effect under the influence of block order. We did observe robust IAT bias in some conditions mentioned above, indicating that MAD and AD were reliable indicators of the IAT effect.

Previous studies showed that word-IAT presented a more substantial IAT effect than picture-IAT (Dasgupta et al., 2000; Foroni & Bel-Bahar, 2010). This study found similar results that participants responded faster when viewing picture stimuli (see Table 2). However, the mouse indices showed a different pattern (see Table 2). For MAD, picture stimuli in the incongruent block first and merged groups showed a significant IAT effect. For AD, trials with picture stimuli in all groups had a significant IAT effect. For AUC, only picture stimuli in the incongruent block first group had the IAT effect. These results indicate that the stimuli modality is more influential on RTs than on mouse dynamic indices. A possible explanation for the modality effect on RTs is that processing an object picture (perceptual memory) requires less cognitive load than processing a sentence or phrase (semantic and situational memory), which therefore reduces the task difficulties (Dasgupta et al., 2000). There is an indication that as the difficulty of the IAT task decreases, so does the magnitude of the IAT effect (Dasgupta et al., 2000). However, the stimulus modality effect on mouse-tracking indices was relatively weak. It seems that the stimulus modality only alters the cognitive complexity of the task without influencing the general effect of implicit memory on individual’s other behavioral performance. Being the initial attempt of applying different stimulus modalities and mouse-tracking techniques in aIAT task, the present results alone cannot draw a general conclusion on how stimulus modality affects MT-aIAT. Further autobiographical-event-relevant studies with more focus on the effect of stimulus modality integrating mouse-tracking are therefore suggested.

#### *4.3. How Can New Computational Modeling Approaches Help*

In MT-aIAT, When participants perceived a stimulus, the stimulus automatically activates the concept itself as well as spreading activation to its linked

associations. The strength of the associations varies between congruent and incongruent blocks. The well-documented finding in the literature is that activation spreads faster if the association between concepts is strong, and spreads more slowly when the strength of association is weak (Verhulst & Lodge, 2013). Meanwhile, cues of prior experience can then trigger an essentially ‘automatic’ pattern of activation in memory that can be described in neural network or connectionist models (Hopfield & Tank, 1986; Queller & Smith, 2002). So according to the analogous features of MT-aIAT and the connectionist model, we hypothesize that the existence of objective descriptions can be disturbed or modified by input of recalling of personal experience during MT-aIAT. Following previous work using output from the self-image network to drive a DDM of binary decision making (Bedder et al., 2019; Wong & Wang, 2006; Wong et al., 2007), we can also infer the magnitude and reactivation of the memory during the response.

In our paper, we postulate a simple connectionist model to quantify the strengths of sensory evidence as well as their associations and use the DDM to model behavioral performance after perceiving sensory information. Specifically, we use the sensory evidence derived from the connectionist model to drive the DDM, measuring the extent to which the extracted information is integrated into choice processes by measuring the coordinates along the cursor trajectory. Additionally, we obtain the simulated RTs from the DDM to further observe the IAT effect. In sum, both our model and the traditional DDM can capture the characteristics of IAT. For the future work, our attempts can also be used in other similar paradigms that follow, such as aIAT and IAT in other contexts, in combination with mouse tracking.

Connectionist models have a profound impact in cognitive science, particularly in the research of memory and perception. This framework assumes that cognitive systems are information processing systems that take in external sensory to form internal representations of the environment, based on which people’s overt behaviors are generated. This framework codes autobiographical events is likely to have been identified by neuroimaging studies which seek the



neural correlates of self-recognition (Feinberg & Keenan, 2005; Keenan et al., 2001). We presented that the connectionist model could offer a mechanistic account of resonance with crime-related events, and explain mental associations and their influence on the aIAT performance. Also, the temporal analysis of mouse trajectories offered a deeper insight into decision dynamics about how much effect the stimuli had on the mouse trajectories. It presented a mechanistic account for implicit bias: it centered on a connectionist network (Hopfield, 1982) composed of neurons responding selectively to different features of stimuli and their attributes. Once this network learns associations between autobiographical stimuli and its attribute, the agent will generate recurrent excitation which promotes total network output.

In standard binary decision making tasks, such as the moving dot paradigm (Huk & Shadlen, 2005), sensory evidence in favor of each motor output is experimentally defined (i.e., as the relative proportion of dots moving in each direction). Oppositely, MT-aIAT is not a value-based task, and thus we can not infer the subjective value of each stimulus or precisely define the strength of sensory evidence for each visual stimulus, making it challenging to examine the latent evidence accumulating process in IAT. This recurrent network (Hebb, 1949) offered us a way to quantify the sensory evidence (see equation (8)). It learns through experience and repeated exposure to stimuli and adjusts the connection weights, which echoes the mock crime session before the MT-aIAT task. In this mock crime session, participants gradually established the associative relationships between their attributes and crime events by stealing the credit card. The connection adjustment schemes in recurrent networks are used for processes usually thought of as “memory”. So the recurrent network hypothesized here is maximally active when the agent perceives crime-relevant events, analogous to the reactivation of memory.

By subsequently applying the output of the connectionist model, that is, the reactivated memory, to a drift-diffusion model of perceptual decision-making (Wong et al., 2007; Wong & Wang, 2006), we are able to simulate the behavioral performance on the IAT to measure the quantity of implicit bias. The

drift-diffusion model assumes that a binary decision making process equals the accumulation of two sides of competing evidence. And the decision is made once the one side of evidence reaches the threshold (Ratcliff & Rouder, 2000). So the accumulation speed of the autobiographical association will be faster due to the activation of memory. Consistent with the empirical data, our model illustrated that simulated agents exhibited negative IAT scores.

Note that the inconsistency in the trend of effect size changes across time between the two groups shown in Figure 9 may be due to the block order effect, because these impacts on the mouse trajectory were counterbalanced in merged data as the estimated coefficients did not have any significant difference. We also observed a peak at around 60th time bin in congruent trials of the incongruent block first group. It might be a recurrent resonance of autobiographical events, indicating the existence of implicit bias (Bedder et al., 2019).

#### 4.4. Limitations and Future Directions

As a preliminary attempt of incorporating mouse tracking with aIAT, the current study is subject to limitations. First, the present findings do not support the claim that the MT-aIAT is superior to standard aIAT. To verify the accuracy advantage that mouse tracking brings to aIAT, additional experiments would be desirable. Second, the scope of this study is limited to only one mock crime condition in a laboratory context. Although we used a single experimental scenario setting to better control experimental variables so that factors affecting IAT effects, such as block order, could be effectively investigated, it would also make the generalizability of the findings limited. Third, the current study have only preliminarily explored the implementation of computational modeling in the MT-aIAT method. It did not compare different modeling account of MT-aIAT and perform model comparison.

Taken together, the mouse tracking paradigm still has much room for further exploration and practice in the future, especially in combination with IAT and aIAT. Though some studies have shown high accuracy of the false identity detection task based on mouse-tracking, indicating high reliability and validity

of mouse-tracking-based memory detection in specific contexts (Monaro et al., 2017a,b), further efforts need to be done in order to validate the predictive accuracy of MT-aIAT and refine the compatibility of mouse-tracking and aIAT in various contexts. As mouse-tracking is covert that the test can be administered to the subject without revealing its memory detection purpose (Sartori et al., 2018), researchers should be aware of the massive ethical implications of implementing a mouse-tracking-based memory detection task and give informed consent to the subjects properly. Besides, a considerable controversy in IAT and aIAT is that discrimination of effect size on the individual level is not strong enough (Blanton et al., 2009; Vargo & Petr  czi, 2013; Greenwald et al., 2015), future efforts on improving the predictive validity can be one step further towards possible practical applications. Also, further research with computational modeling can be developed to reveal the possible cognitive process during aIAT with the information provided by the mouse-tracking technique. Another notable focus of research on the aIAT approach is the neurophysiological basis such that future studies can exploit the investigation IAT brain mechanism as an entry point to advance the exploration of the principles of deception behavior as well as its detection.

To provide some reference for future studies that want to explore and improve MT-aIAT, we list below some suggestions.

- (1) An immersive instruction or cover story. The instruction was verbally transferred by the experimenter and try to increase the participants' sense of involvement.
- (2) Combing with strict scene layout and ecological contexts. To control participants' autobiographical memory, we strictly fixed up the experimental room so that all participants had to experience the identical procedure to get the card. The application of MT-aIAT also calls for more ecological autobiographical memory detection in different contexts.
- (3) Reliable mouse-tracking technique. It is critical to record reliable mouse trajectories in the task and provide a standard trajectory preprocessing

that unified the mouse trajectories temporally and spatially, guaranteeing the robustness of the mouse dynamic analysis.

## Conclusion

In sum, this study assessed and confirmed the validity of MT-aIAT, with mouse-tracking and computational modeling methods to probe into the cognitive mechanisms by the MT-aIAT in a mock crime scenario. The mouse tracking data with the temporal analysis can investigate the discrepancy between different conditions at each time point. Taking the block order of the task and stimulus modality into consideration, our approach suggested that the reaction time, MAD and AD were reliable indices to detect autobiographical event. Furthermore, a connectionist model was used to compute the sensory evidence of each stimulus. We regressed the sensory evidence on the trajectory angle at every time point in order to examine the impacts that the stimulus casts on the decision-making process. Output of the connectionist model was then feed into the DDM to simulate the RT-based aIAT effect, which were consistent with our empirical data.

## Declarations

### *Acknowledgments*

This work is supported by National Natural Science Foundation of China (U1736125) and SRG of the University of Macau. The authors have no conflicts of interest to declare that are relevant to the content of this article. Besides, we would like to thank Tianyu Ma, Miao Chang, and Yutong Guo for their consistent help with the experiment. We would also like to thank all participants that took part in the study and enabled this research to be possible.

### *Ethics Approval*

The experimental protocol of this study was approved by the ethical review committee of the Institute of Psychology, Chinese Academy of Sciences.

### *Consent to Participate*

Informed consent was obtained from all individual participants included in the study.

### *Open Practices Statement*

The data and materials for all experiments are available on GitHub at (<https://github.com/andlab-um/MT-aIAT>), and none of the experiments was preregistered.

### *Availability of Data and Materials*

The authors confirm that the anonymized behavioral data supporting our main analyses are available on GitHub (<https://github.com/andlab-um/MT-aIAT>).

### *Code Availability*

The code that support the findings of this study are openly available in the lab GitHub at (<https://github.com/andlab-um/MT-aIAT>).

### *Author Contributions*

X. Xu and H. Wu conceived the research, H. Wu and X. Xu performed the research, X. Xu and H. Wu analyzed the data, X. Xu, X. Liu, X. Hu and H. Wu wrote the paper.

## **References**

- Agosta, S., Mega, A., & Sartori, G. (2011). Detrimental effects of using negative sentences in the autobiographical iat. *Acta psychologica*, 136, 269–275.
- Agosta, S., & Sartori, G. (2013). The autobiographical iat: a review. *Frontiers in Psychology*, 4, 519.
- Anikin, A., & Johansson, N. (2019). Implicit associations between individual properties of color and sound. *Attention, Perception, & Psychophysics*, 81, 764–777.

- Bedder, R. L., Bush, D., Banakou, D., Peck, T., Slater, M., & Burgess, N. (2019). A mechanistic account of bodily resonance and implicit bias. *Cognition*, *184*, 1–10.
- Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the iat. *Journal of applied Psychology*, *94*, 567.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, *113*, 700.
- Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, *340*, 95–98.
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., De Rosario, H., & De Rosario, M. H. (2018). Package ‘pwr’. *R package version*, *1*.
- Chassot, S., Klöckner, C. A., & Wüstenhagen, R. (2015). Can implicit cognition predict the behavior of professional energy investors? an explorative application of the implicit association test (iat). *Journal of Applied Research in Memory and Cognition*, *4*, 285–293.
- Chemin, B., Huang, G., Mulders, D., & Mouraux, A. (2018). Eeg time-warping to study non-strictly-periodic eeg signals related to the production of rhythmic movements. *Journal of neuroscience methods*, *308*, 106–115.
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & cognition*, *35*, 15–28.
- Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for white americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*, 316–328.

- Dong, G., Wu, H., & Lu, Q. (2010). Attempting to hide our real thoughts: Electrophysiological evidence from truthful and deceptive responses during evaluation. *Neuroscience letters*, *479*, 1–5.
- Duran, N. D., Dale, R., & McNamara, D. S. (2010). The action dynamics of overcoming the truth. *Psychonomic bulletin & review*, *17*, 486–491.
- Feinberg, T. E., & Keenan, J. P. (2005). Where in the brain is the self? *Consciousness and cognition*, *14*, 661–678.
- Feroni, F., & Bel-Bahar, T. (2010). Picture-iat versus word-iat: level of stimulus representation influences on the iat. *European Journal of Social Psychology*, *40*, 321–337.
- Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in psychology*, *2*, 59.
- Freeman, J. B., & Ambady, N. (2010). Mousetracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior research methods*, *42*, 226–241.
- Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*, 673.
- Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, *46*, 179–185.
- Gonick, L., Smith, W., & Smith, W. (1993). *The cartoon guide to statistics*. HarperPerennial New York.
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the implicit association test can have societally large effects., .

- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74, 1464.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of personality and social psychology*, 85, 197.
- Hebb, D. O. (1949). The organization of behavior; a neuropsychological theory. *A Wiley Book in Clinical Psychology*, 62, 78.
- Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology Press.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79, 2554–2558.
- Hopfield, J. J., & Tank, D. W. (1986). Computing with neural circuits: A model. *Science*, 233, 625–633.
- Hu, X., Rosenfeld, J. P., & Bodenhausen, G. V. (2012). Combating automatic autobiographical associations: The effect of instruction and training in strategically concealing information in the autobiographical implicit association test. *Psychological Science*, 23, 1079–1085.
- Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience*, 25, 10420–10436.
- Keenan, J., McCutcheon, N., & Pascual-Leone, A. (2001). 18. functional magnetic resonance imaging and event-related potentials suggest right prefrontal activation for self-related processing. *Brain and Cognition*, 47, 87–91.
- Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior research methods*, 49, 1652–1667.



- Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the implicit association test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, 93, 353.
- Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive psychology*, 67, 151–185.
- Marini, M., Agosta, S., & Sartori, G. (2016). Electrophysiological correlates of the autobiographical implicit association test (aiat): response conflict and conflict resolution. *Frontiers in human neuroscience*, 10, 391.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, 132, 455.
- Monaro, M., Bertomeu, C. B., Zecchinato, F., Fietta, V., Sartori, G., & Martínez, H. D. R. (2021a). The detection of malingering in whiplash-related injuries: a targeted literature review of the available strategies. *International journal of legal medicine*, (pp. 1–16).
- Monaro, M., Gamberini, L., & Sartori, G. (2017a). The detection of faked identity using unexpected questions and mouse dynamics. *PloS one*, 12, e0177851.
- Monaro, M., Gamberini, L., & Sartori, G. (2017b). Identity verification using a kinematic memory detection technique. In *Advances in neuroergonomics and cognitive engineering* (pp. 123–132). Springer.
- Monaro, M., Negri, P., Zecchinato, F., Gamberini, L., & Sartori, G. (2021b). Mouse tracking iat in customer research: An investigation of users’ implicit attitudes towards social networks. In *International Conference on Intelligent Human Systems Integration* (pp. 691–696). Springer.
- Nahari, G. (2018). The applicability of the verifiability approach to the real world. In *Detecting concealed information and deception* (pp. 329–349). Elsevier.

- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, 6, 101.
- Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic bulletin & review*, 19, 906–913.
- Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic bulletin & review*, 24, 1234–1251.
- Queller, S., & Smith, E. R. (2002). Subtyping versus bookkeeping in stereotype learning and change: Connectionist simulations and empirical findings. *Journal of personality and social psychology*, 82, 300.
- Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human perception and performance*, 26, 127.
- Rosenfeld, J. P. (2018). Detecting concealed information and deception: Recent developments, .
- Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological science*, 19, 772–780.
- Sartori, G., Zangrossi, A., & Monaro, M. (2018). Deception detection with behavioral methods: the autobiographical implicit association test, concealed information test–reaction time, mouse dynamics, and keystroke dynamics. In *Detecting Concealed Information and Deception* (pp. 215–241). Elsevier.
- Smeding, A., Quinton, J.-C., Lauer, K., Barca, L., & Pezzulo, G. (2016). Tracking and simulating dynamics of implicit stereotypes: A situated social cognition perspective. *Journal of Personality and Social Psychology*, 111, 817.

- Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, 15, 207–211.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102, 10393–10398.
- Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological science*, 26, 122–134.
- Van Ravenzwaaij, D., van der Maas, H. L., & Wagenmakers, E.-J. (2011). Does the name-race implicit association test measure racial prejudice? *Experimental Psychology*, 58, 271.
- Vargo, E. J., & Petróczi, A. (2013). Detecting cocaine use? the autobiographical implicit association test (aiat) produces false positives in a real-world setting. *Substance abuse treatment, prevention, and policy*, 8, 1–13.
- Verhulst, B., & Lodge, M. (2013). Reaction time measures in implicit attitudes research. In *Political science research methods in action* (pp. 64–92). Springer.
- Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge University Press.
- Verschuere, B., Prati, V., & Houwer, J. D. (2009). Cheating the lie detector: Faking in the autobiographical implicit association test. *Psychological Science*, 20, 410–413.
- Vrij, A., & Fisher, R. P. (2016). Which lie detection tools are ready for use in the criminal justice system? *Journal of Applied Research in Memory and Cognition*, 5, 302–307.
- Wong, K.-F., Huk, A. C., Shadlen, M. N., & Wang, X.-J. (2007). Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Frontiers in Computational Neuroscience*, 1, 6.

- Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, *26*, 1314–1328.
- Wu, H., Hu, X., & Fu, G. (2009). Does willingness affect the n2-p3 effect of deceptive and honest responses? *Neuroscience letters*, *467*, 63–66.
- Yu, Z., Wang, F., Wang, D., & Bastin, M. (2012). Beyond reaction times: Incorporating mouse-tracking measures into the implicit association test to examine its underlying process. *Social Cognition*, *30*, 289–306.

## Tables and Figures

Table 1: The congruency effect for different data.

Index	P value of Data		
	Congruent Block First	Incongruent Block First	Merged
RT	0.041 *	<0.001 ***	<0.001 ***
MAD	.055	.0083**	0.013*
AD	.026*	.0071**	<0.001***
AUC	.16	.037 *	.012*

*Note.* P-values of paired t-tests of different indices between congruent and incongruent trials. The results showed that IAT bias existed not only in reaction times but also mouse geometric indices.\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 2: t tests between congruent and incongruent block of picture and word stimuli in each group.

Index	P value of Data		
	Congruent Block First	Incongruent Block First	Merged
RT-picture	t = -1.88, p = 0.069	t = -5.32, p < 0.001 ***	t = -4.74, p < 0.001 ***
RT-word	t = -3.16, p = 0.0035 **	t = -8.47, p < 0.001 ***	t = -6.95, p < 0.001 ***
MAD-picture	t = -6.95, p < 0.001 ***	t = -3.46, p = 0.0022 **	t = -3.68, p < 0.001 ***
MAD-word	t = 0.60, p = 0.55	t = -0.65, p = 0.52	t = 0.01, p = 0.99
AD-picture	t = -2.19, p = 0.037*	t = -3.30, p = 0.0033 **	t = -3.71, p < 0.001 ***
AD-word	t = 0.41, p = 0.69	t = -0.89, p = 0.38	t = -0.26, p = 0.79
AUC-picture	t = 0.13, p = 0.90	t = -2.64, p = 0.015 *	t = -1.80, p = 0.078
AUC-word	t = 0.96, p = 0.34	t = 0.054, p = 0.96	t = 0.77, p = 0.44

*Note.* For RT, word stimuli had a stronger IAT effect than picture stimuli. For MAD, picture stimuli in incongruent block first group and merged group had significant IAT effect. For AD, trials with picture stimuli in all groups had significant IAT effect. For AUC, only picture stimuli in incongruent block first group had IAT effect.\*  $p < 0.05$ ,

\*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

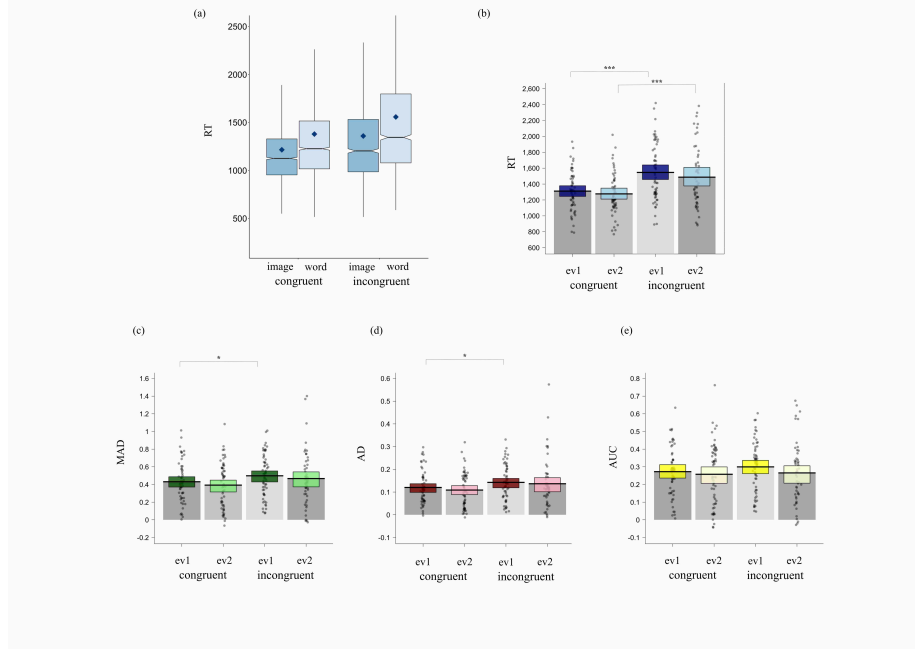


Figure 1: Behavioral results and mouse geometric indices of the merged data. (a) Mean reaction times of word and picture stimuli under different conditions. (b) Mean reaction times of autobiographical and irrelevant event under congruent and incongruent blocks. Significant IAT bias can be observed in both events. (c-e) Mean MAD, AD, AUC of autobiographical and irrelevant event under congruent and incongruent blocks. For MAD and AD, significant IAT biases were observed in autobiographical event. For AUC, we did not observe any discrepancy between two conditions.

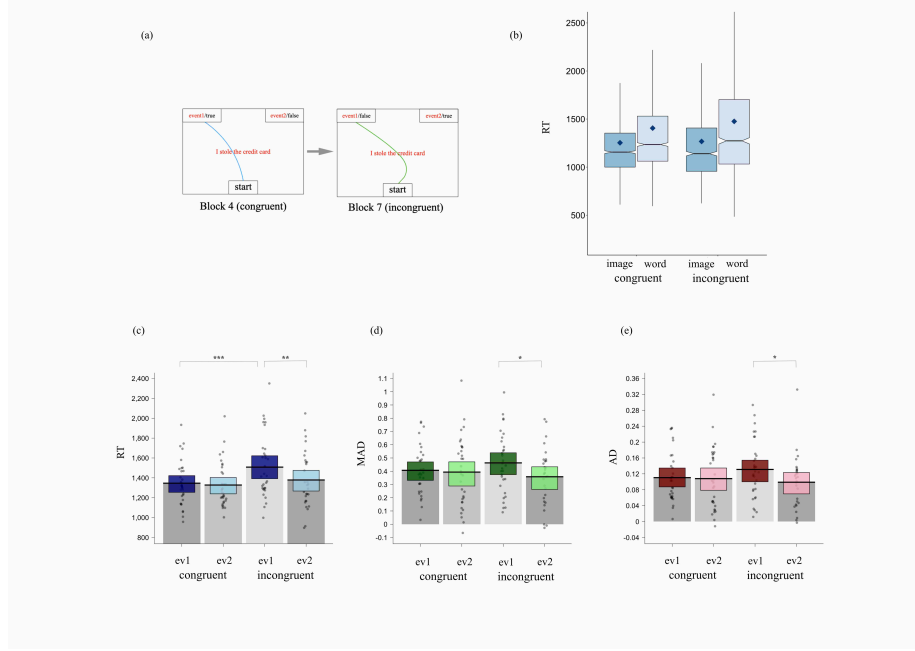


Figure 2: Experiment paradigm, behavioral results and mouse geometric indices of the congruent block first group. (a) Experiment paradigm of the congruent block first group: congruent block came first. (b) Mean reaction times of word and picture stimuli under different conditions. (c) Mean reaction times of autobiographical and irrelevant event under congruent and incongruent blocks. Significant IAT bias can be observed in both events. (d-e) Mean MAD and AD of autobiographical and irrelevant event under congruent and incongruent blocks. Significant IAT bias can be observed in the incongruent condition. Besides, we did not observe any significant discrepancy with AUC.



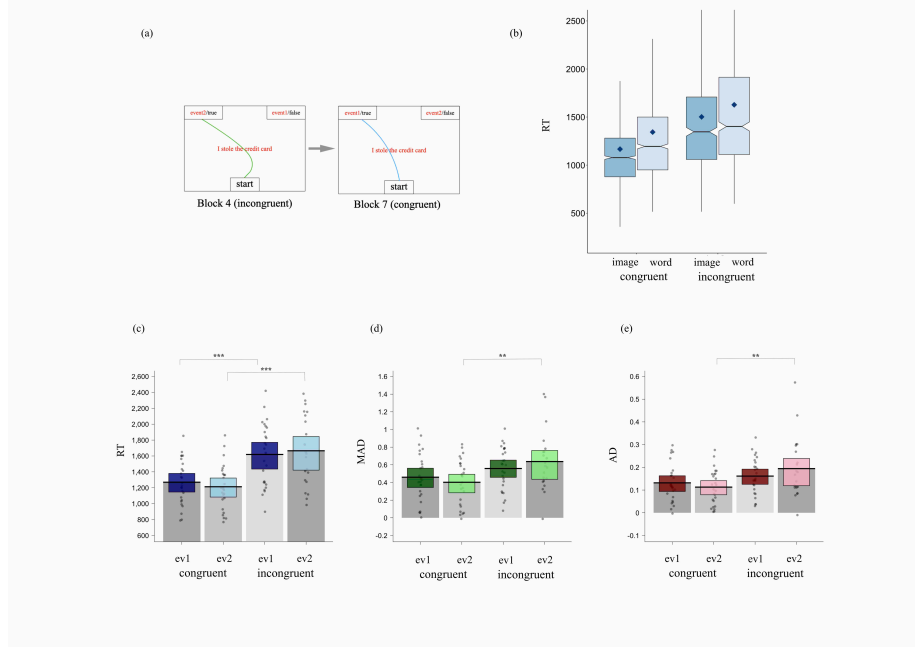


Figure 3: Experiment paradigm, behavioral results and mouse geometric indices of the incongruent block first group. (a) Experiment paradigm of the congruent block first group: incongruent block came first. (b) Mean reaction times of word and picture stimuli under different conditions. (c) Mean reaction times of autobiographical and irrelevant event under congruent and incongruent blocks. Significant IAT bias can be observed in both events. (d-e) Mean MAD and AD of autobiographical and irrelevant event under congruent and incongruent blocks. No significant IAT biases were observed in autobiographical event. Only the difference of irrelevant events between in congruent and incongruent condition was significant. Besides, we did not observe any significant discrepancy with AUC.

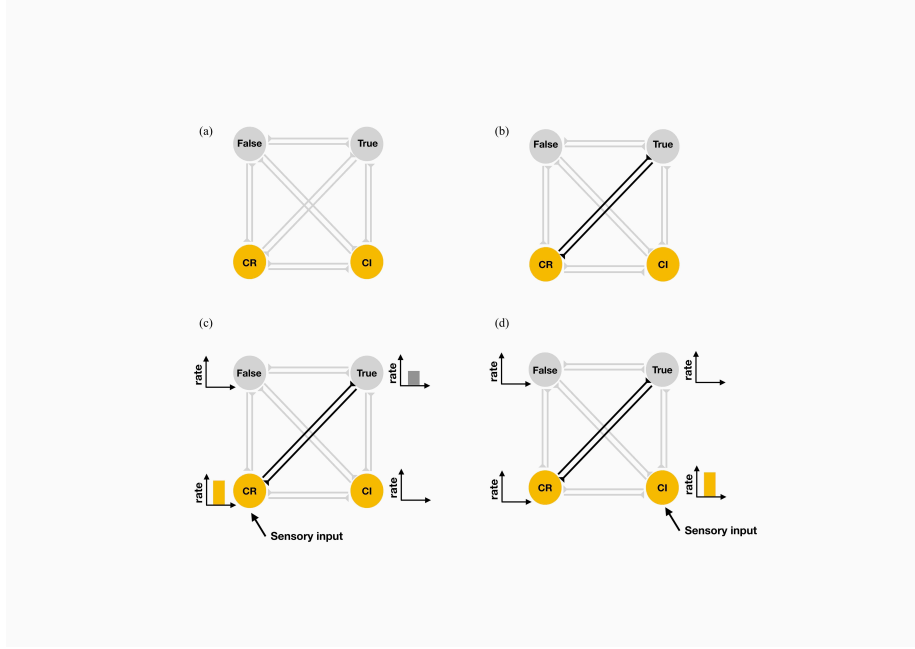


Figure 4: (a) The structure of Connectionist Model. (b) The simulated agent committed crime-relevant events as our participants did, such that the connectionist model comes to encode strong associations between neurons encoding true and crime-relevant features. (c, d) During subsequent perception, sensory input to sub-populations encoding the autobiographical feature of that agent generates additional activity via recurrent synaptic connections if those features overlap with the encoded feature.

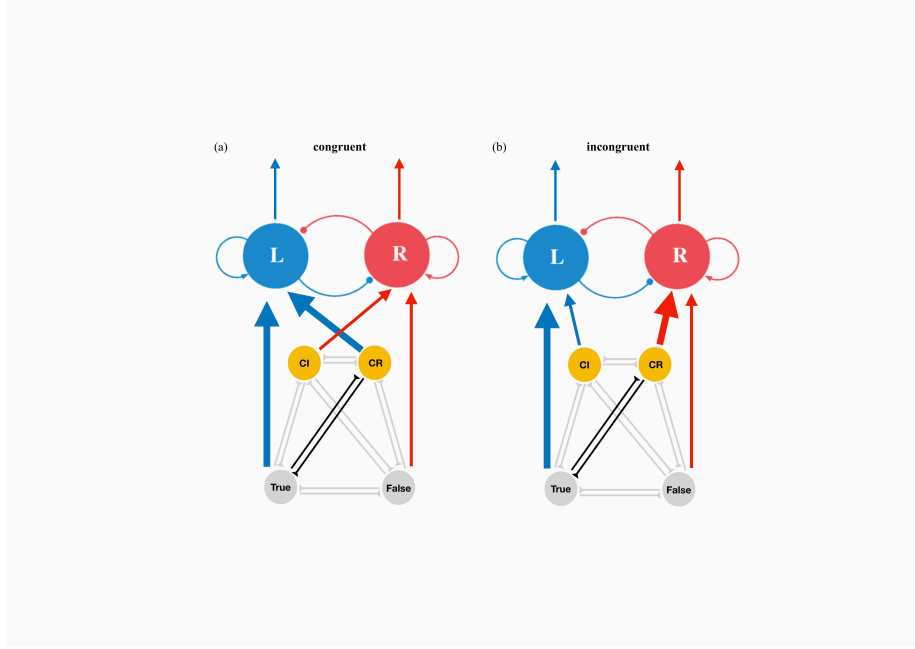


Figure 5: The schematic overview of the drift diffusion model (DDM). The DDM consists of two self- excitatory but mutually inhibitory neural populations coding for left and right motor outputs. In our simulations, the sensory evidence provided to each DDM motor population in each IAT trial is determined by activity in the connectionist model. Neurons coding for the IAT stimulus receive a set level of external sensory input, while additional input to either motor response population arises from recurrent excitation within the connectionist model (indicated by thicker colored arrows). Note that connections from the connectionist model to the DDM are flexibly reconfigured between congruent and incongruent blocks.

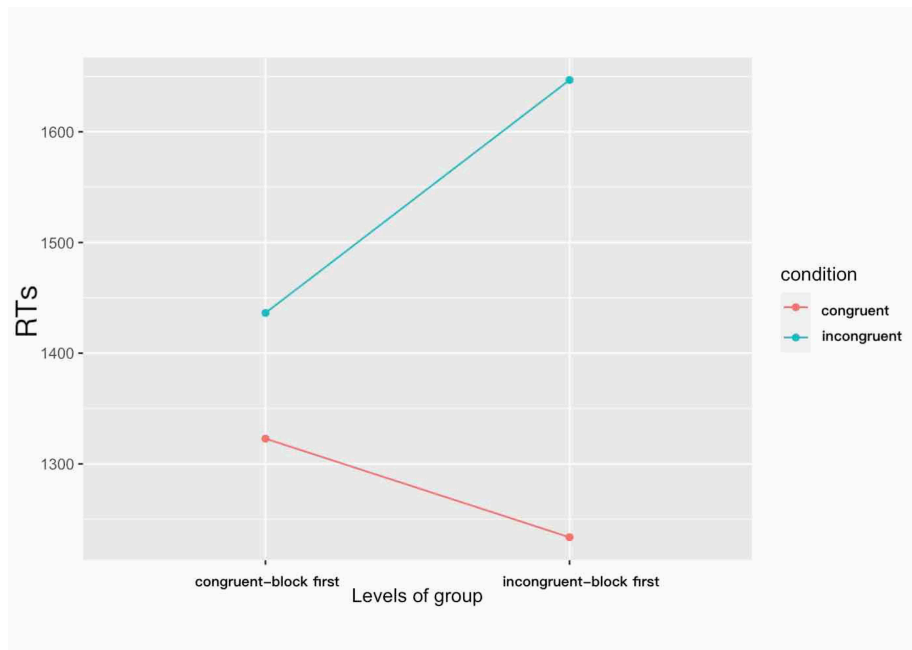


Figure 6: Interaction between the order and conditions. Both orders showed strong IAT effect (congruent-block first:  $t = -3.17, p = 0.0027$ ; incongruent-block first:  $t = -10.00, p < 0.001$ ), and the incongruent-block first group had a significantly bigger IAT effect ( $t = 5.48, p < 0.001$ ).

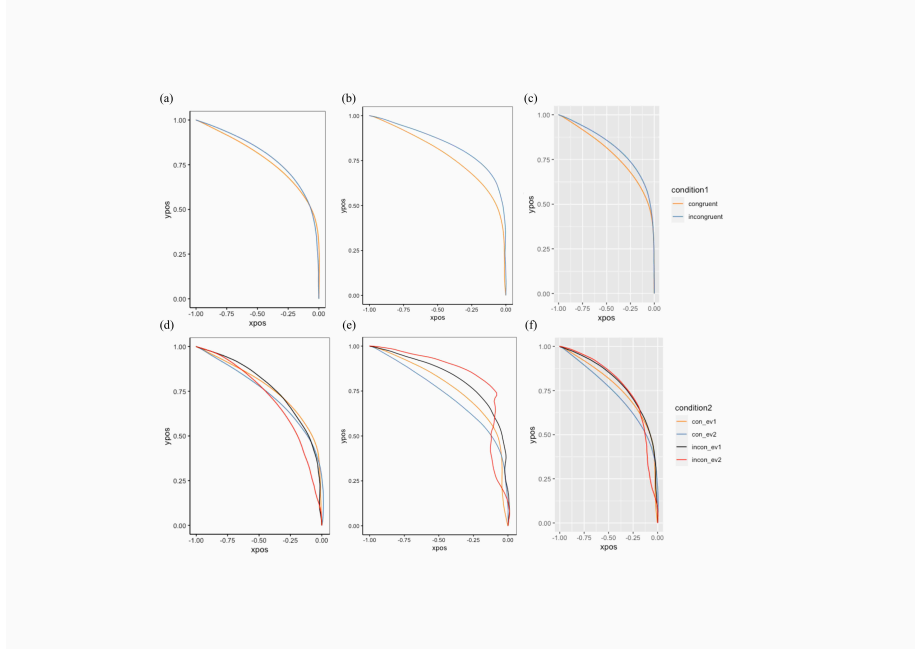


Figure 7: (a-c) Aggregated mouse trajectories of the congruent block first group, the incongruent block first group and the merged data under congruent and incongruent conditions. All trajectories were remapped leftward. (d-f) Aggregated mouse trajectories of the congruent block first group, the incongruent block first group and the merged data when viewing event1 and event2 under congruent and incongruent conditions. All trajectories were remapped leftward.

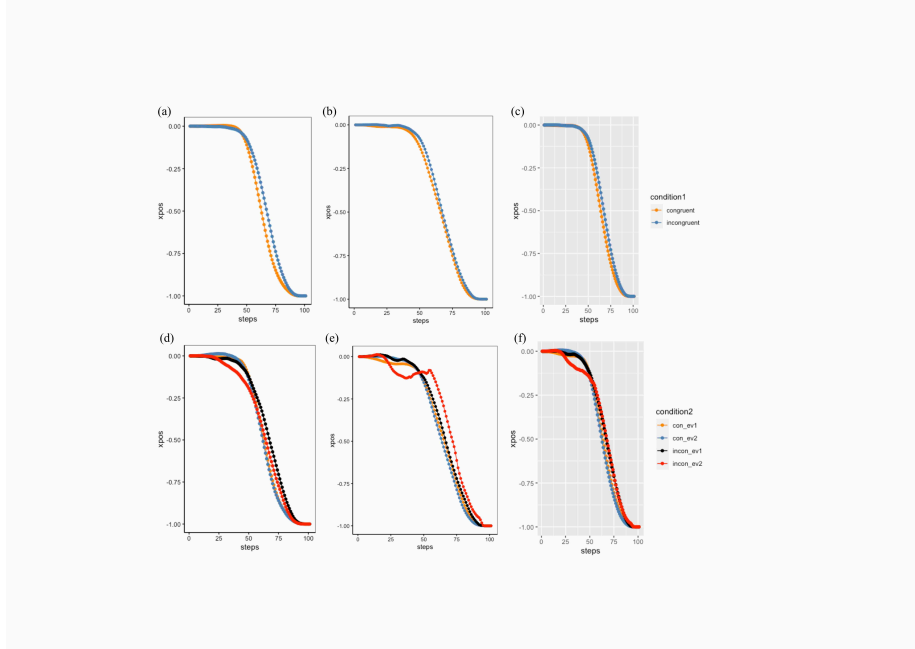


Figure 8: (a-c) Aggregated mouse trajectories of the congruent block first group, the incongruent block first group and the merged data under congruent and incongruent conditions. All trajectories were remapped leftward. (d-f) Aggregated mouse trajectories of the congruent block first group, the incongruent block first group and the merged data when viewing event1 and event2 under congruent and incongruent conditions. All trajectories were remapped leftward.

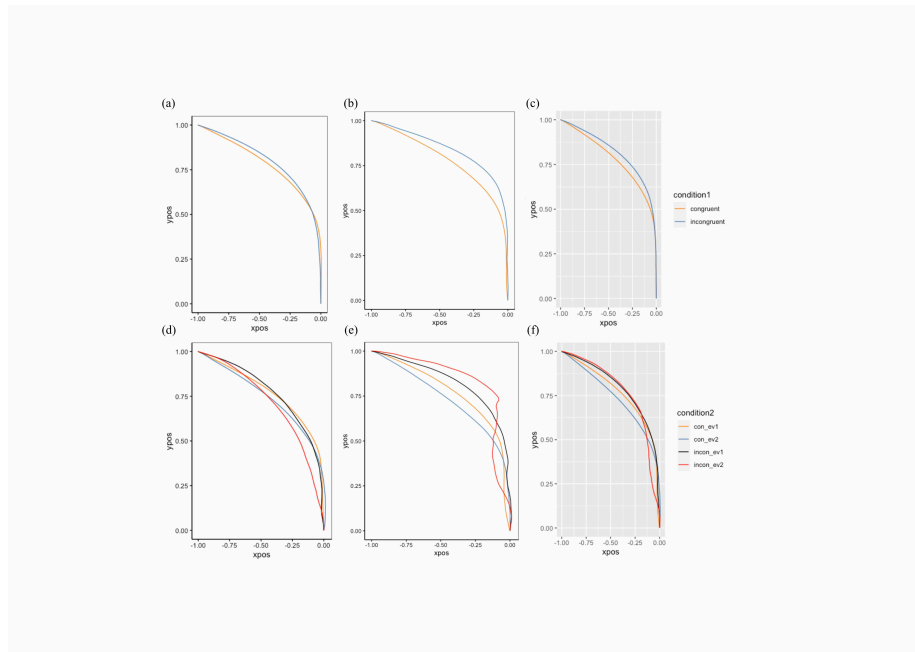


Figure 9: The congruent block first group, the incongruent block first group, merged data: Estimated coefficients of linear regression that examined how the trajectory angle was dependent on each stimulus of different condition in every normalized time point.

## Supplementary Materials 1

Experimental stimuli used in aIAT task:

### Word stimuli

#### I. Logically -true or -false events

Logically-true		Logically-false	
<i>Chinese</i>	<i>English</i>	<i>Chinese</i>	<i>English</i>
我在北京	I am in Beijing.	我在国外	I am abroad.
我是女生	I am a female.	我是男生	I am a male.
我在三楼	I am at the third floor.	我在楼顶	I am at the top of the building.
我在房内	I am in a room.	我在商场	I am in a market.
我在楼里	I am in a building	我在浴室	I am in the bathroom.

#### II. Crime -relevant or -irrelevant events

Crime-relevant		Crime-irrelevant	
<i>Chinese</i>	<i>English</i>	<i>Chinese</i>	<i>English</i>
找到钥匙	I found the key.	插入 u 盘	I inserted the USB.
拿银行卡	I took the credit card.	找到插口	I found the USB port.
推开凳子	I pushed aside the chair.	找到密码	I found the keyword.
打开抽屉	I opened the drawer.	撕开信封	I opened the envelope.
翻开钱包	I opened the wallet.	复制文件	I copied the file.



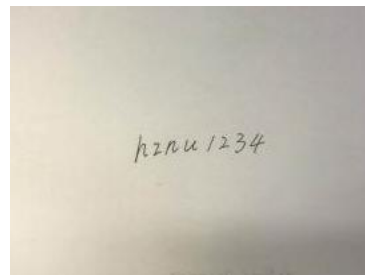
Picture stimuli

Crime -relevant or -irrelevant events

Crime-relevant



Crime-irrelevant



## Supplementary materials 2

The present study used MATLAB for the simulations of RTs. The connectionist model used in the present work is a simple recurrent neural network (RNN), consisting of four artificial neurons that represent four features (the crime-relevant event, the crime-irrelevant event, true, and false). The link strengths among the neurons are the connection weights learned based on the Hebbian learning rule [? ]. Assuming a fully-connected recurrent network, the firing rate output  $r_i$  of neuron  $i$  is controlled by a linear activation function which transforms total input to a firing rate output, with a peak firing rate of  $r_{max} = 10$  Hz and a time constant of  $\tau_r = 10$  ms (Eqs. (1) and (2)). The total input to each neuron ( $I_{tot,i}$ ) is a sum of external sensory input ( $I_{ext,i} = 0.5$ ) and recurrent synaptic currents ( $I_{rec,i}$ ) (Eq. (3)). External sensory inputs occurs when the agent received the corresponding stimulus. As is shown in Eq.(4), recurrent synaptic currents are defined as the product of firing rates of connected neurons  $r_j$  and synaptic weights  $w_{ij}$ . These recurrent synaptic currents contribute to the variations in the synaptic connection strength; that is, the updated weights are the sum of pre-updated weights and recurrent synaptic currents (Eq.(5)). Here we set the learning rate  $k = 2 \times 10^5$ . All synaptic connections  $w_{ij}$  and firing rates  $r_i$  are initially set to zero. Besides, we presume an abstract neuromodulatory signal  $e$  that distinguishes between periods of encoding ( $e = 1$ ) and retrieval ( $e = 0$ ). In the encoding period, the network builds associations between features belonging to the agent(the crime-relevant event true) and the recurrent synaptic currents do not exist. During the period of retrieval, recurrent synaptic currents are activated and passed to the next neuron as input when the agent receives the features same to their own.

This dynamic system will finally reach a stable state (Eq.(2)), with firing rates as the input of the drift-diffusion model.

$$I_{tot,i} = \begin{cases} I_{tot,i} & \text{for } I_{tot,i} \leq r_{max} \\ r_{max} & \text{for } I_{tot,i} > r_{max} \end{cases} \quad (1)$$

$$\tau_r \frac{dr_i}{dt} = -r_i + I_{tot,i} \quad (2)$$

$$I_{tot,i} = I_{ext,i} + I_{rec,i} \quad (3)$$

$$I_{rec,i} = (1 - e) \sum_{n=1}^j r_n w_{in} \quad (4)$$

$$w_{ij} = w_{ij} + e k r_i r_j \quad (5)$$

The drift-diffusion models (DDM) are established based on the two-alternative forced choice tasks. In the DDM, two competing neuronal pools accumulate evidence of motor responses on the left and right, respectively. When evidence in one of the populations reaches the pre-determined firing rate threshold, the motor response is enacted and the decision is made. In this study, we apply the two-variable network model reported by [?] and [?](see Figure 6(a)). We obtained the stimuli-driven input of the DDM from the connectionist model once firing rates of it reach an equilibrium condition. We acquired each neuron's firing rate output from the connectionist model and subsequently calculated the total firing rate output of each one of the motor populations. Then the level of sensory evidence  $s'$  is proportional to the difference in total firing rate input for each motor population, not including the firing rate activity produced by external stimulation  $I_{ext}$  (Eq. (8)). We made use of the sensory evidence  $s'$  [?] to gain the input  $I_{stim}$  caused by stimuli for each neuron population, which is proportional to the sensory evidence  $s'$  adjusted by a gain factor  $f = 0.45$  and a mean value of  $0 = 30$  Hz (Eq. (7)).

$$I_{fix} = \begin{cases} J_{A,ext}(50 + 100\exp[-(t - t_{fix})/\tau_{ad}]) & \text{for } t_{fix} \leq t < t_{stim} \\ J_{A,ext}(6 + 44\exp[-(t - t_{stim})/\tau_{ad}]) & \text{for } t > t_{stim} \end{cases} \quad (6)$$

$$I_{stim,i} = J_{A,ext}\mu_0(1 + f s') \quad (7)$$

$$s' = (\sum r_L - \sum r_R) - I_{ext} + 0.05 \quad (8)$$

Fixation on the cross before every trial contributes to a small excitatory input  $I_{fix}$  defined in Eq.(6). It is adjusted by a gain factor  $J_{A,ext} = 1.1 \times 10^3$  nA/Hz and shows an adaptation with a time constant  $ad = 40$  ms. Each population also receives noisy synaptic inputs with a mean value of  $I_0 = 0.3297$  nA and a white noise component  $\eta_i(t)$  with an amplitude of  $\sigma_{noise} = 0.009$  nA filtered by a synaptic time constant of  $\tau_{noise} = 2$  ms (Eq. (12)). Each population's firing rate  $r_i$  (where  $i = L, R$ ) is a function of the synaptic current input  $I_i$  described in Eq.(2a) [see ? ], with parameters  $a = 270$  Hz/nA,  $b = 108$  Hz and  $d = 0.154$  s.

Together with stimuli-driven input, fixation excitatory input and noise, the excitatory and inhibitory synaptic couplings constitute the total synaptic input to each population with  $J_{ii} = 0.3275$  nA and  $J_{ij} = 0.1137$  nA (Eq. (10)). NMDA currents  $S_i$  are manipulated by a dynamic system described in Eq. (11).

At last, we obtained the lower firing rate of two populations as the simulated reaction time through Eq(9) with  $a = 270$  Hz/nA,  $b = 108$  Hz and  $d = 0.154$  s.

$$r_i = f(I_i) = \frac{aI_i - b}{(1 - \exp[-d(aI_i - b)])} \quad (9)$$

$$I_{tot,i} = J_{ii}S_i - J_{ij}S_j + I_{stim,i} + I_{fix} + I_{noise,i} \quad (10)$$

$$\frac{dS_i}{dt} = -\frac{S_i}{\tau_s} + (1 - S_i)\gamma f(I_i) \quad (11)$$

$$\tau_{noise} \frac{dI_{noise}(t)}{dt} = -(I_{noise,i}(t) - I_0) + \eta_i(t)\sqrt{\tau_{noise}}\sigma_{noise} \quad (12)$$