Transferring Cognitive Talent Across Domains: The Case of Finance

K. Rotaru, P. Kalev, N. Yadav, P. Bossaerts

SUPPLEMENTARY INFORMATION

1. Experimental Design in Detail

1.1. Summary of procedures

Our training protocol is organized as follows. There are two interventions, separated by four weeks to allow for treatment washout and subsequent re-uptake. Each of the interventions consists of four sessions: (i) a session organized along the lines of Frydman et al. (2014), to measure the extent to which the participant's investment decisions are affected by the disposition effect; (ii) a session whereby the participant (advisor) chooses investments for a person (advisee) selected by the advisor; importantly, the advisee *always realizes* gains (or losses) immediately after the end of an investment trial, so the advisee is never invested when the participant recommends investment; (iii) a session that combines (i) and (ii), whereby the advisor chooses investments for herself, as well as for the advisee (selected in the previous session) who immediately realizes gains (or losses) upon conclusion of an investment trial; (iv) a repetition of session (i), to measure the impact on the disposition effect of the training intervention.

The investment game is borrowed from Frydman et al. (2014), but is limited to one security rather than three. In addition, participants can change investments after each trial. As such, it is easier to implement the Bayes-optimal strategy.

The situation is simple. Participants take positions (long; short) in one share of a stock that goes through good and bad "regimes." In the good regime, the stock price goes up the majority of the time; in the bad regime, the stock mostly goes down. Regime switches happen randomly. Participants know that there are regime switches, which occur randomly over time, as well as the possible magnitudes of the outcomes in each regime.

We measure the disposition effect as in Odean (1998). The measure penalizes for paper losses and realized gains, since the Bayes-optimal policy is to stay invested upon gains (suggesting that the stock is in the good regime), while divesting, and even shorting, upon losses (which indicate that the stock is in the bad regime).

1.2. Participants

Seventy six undergraduate and postgraduate students voluntarily participated in this study. Eight participants (10.53% of the original sample) were removed from the sample as they did not attend the retest laboratory sessions. The behavioral analyses were therefore based on the remaining 68 participants. 52.9% of the remaining participants were male, 45.6% were female, while one participant (1.5% of the sample) preferred not to disclose the gender. The average age of the participants was 21.22 years (*SD* = 1.92), ranging from 18 to 27 years. 16.2% of the participants had prior trading experience outside the academic setting, while the rest (83.8%) did not have such experience. In the analyses of eye tracking data, further five participants had to be removed due to a high amount of missing eye tracking data[1] (we imposed a data quality threshold of 85%; see Appendix A).

The choice of sample size was based on the study of Frydman and Rangel (2014). By reducing the visual saliency of the purchase price, Frydman and Rangel found a significant reduction in the disposition effect with 58 participants. As such, a sample size of 58 appears to generate sufficient power to evaluate the efficacy of interventions. Consequently, we aimed at the same number of participants. We ended up with more valid records (68 instead of 58). As we shall see, we too find a significant effect. Importantly, our effect size will be far bigger.

1.3. Intervention Design

In this experimental study, we consider a longitudinal[2] pre-post intervention design (e.g. Harris et al., 2006). We aim to investigate the effectiveness of the proposed ToM-based training scheme to reduce the disposition effect. To test for potential washout of the treatment effect, the experimental

---

[1] The exclusion rate due to missing eye tracking data reported in our study is in line with prior eye tracking literature (e.g., van Rijn, Dalenburg, Borst, & Sprenger 2012; Chiew & Braver, 2013).
[2] This means that participants had to make a sequence of decisions across multiple trading sessions.

treatments are administered twice (test and retest sittings) within an interval of four weeks. The disposition effect is measured at the beginning and end of each sitting.

In the statistical analysis of the results, we choose the first set of *dependent variables (DVs)* to be based on an individual measure of the disposition effect, as operationalized in Odean (1998) and utilized among others by Frydman et al. (2014), Frydman and Rangel (2014), Goulart et al. (2015), and Fischbacher et al. (2017). This measure is calculated as the difference between the Proportion of Gains Realized (PGR) and Proportion of Losses Realized (PLR). PGR is calculated as a ratio between the number of realized gains (market price of the stock is above the purchase price, and the participant decides to sell the stock) and the sum of the number of realized gains and the number of paper gains (market price of the stock is above the purchase price, but the participant decides not to sell the stock). PLR is calculated as the ratio between the number of realized losses (market price of the stock is below the purchase price, and the participant decides to sell the stock) and the sum of the number of realized losses and the number of paper losses (market price of the stock is below the purchase price, but the participant decides not to sell the stock).

The second set of *DVs* is based on the *difference* between individual disposition effect scores obtained for each of the trading sessions. They are meant to provide a measure of learning and improvement.

The third set of *DVs* is associated with the degree of a participant's attention to the acquisition price compared to overall attention paid to the trading dashboard. These measures capture the proportion of eye fixations on the acquisition price relative to eye fixations on the overall dashboard.

Independent variables (IVs) are based on ToM, as assessed using three subscales of the Awareness of Social Inference Test - Revised (TASIT-R; Flanagan et al., 2011), which delineate and measure individual neurocognitive skills associated with social and emotional cognition. The subscales are: the Emotion Evaluation Test (EET), the Social Inference–Minimal (SI-M) test, and the Social Inference–Enriched (SI-E) test. With these IVs, we investigate whether the level of social cognition (as per the social inference subscales of the TASIT-R test) versus the level of emotional cognition (as per the emotional inference subscale of the TASIT-R) are associated with (i) the *level* of the disposition effect, and (ii) the *reduction* of the disposition effect before and after the experimental interventions.

1.4. Stock Trading Task

The experimental task closely follows that of Frydman, et al. (2014), Frydman and Rangel (2014), which is based on the stock trading task introduced by Weber and Camerer (1998). Participants are given the opportunity to trade one stock, named stock A. The experiment is based on four sessions separated by a two-minute break. Each session lasts for an average of 17 minutes and consists of 100 trials. Before each session, the participant is given $50 in experimental currency. The participant is then asked to buy one share of stock. The initial share price for stock A is $100. Cash positions could become negative. No interest is charged on negative cash positions.

Each participant is allowed to hold a maximum of one ('1') share and a minimum of minus one ('-1') shares (negative positions correspond to short-selling). The participants can also hold zero ('0') shares. The price at which a participant can buy or sell is given by the current market price. Once a stock is purchased, the acquisition price is updated on the trading dashboard (see Figure 1).

Figure 1. Trading dashboard.
Version of the research instrument where participants traded for their own account.

The price path of stock A is governed by a two-state Markov chain with a good state and a bad state. Suppose that, in trial t, where $t \in \{1,2,\ldots,100\}$, there is a price update for stock A. If stock A is in the good state at that time, its price increases with probability 0.55 and decreases with probability 0.45. Conversely, if it is in the bad state at that time, its price increases with probability 0.45 and decreases with probability 0.55. The magnitude of the price change is drawn uniformly from {\$5, \$10, \$15}, independently of the direction of the price change.

The state of each stock evolves over time in the following way. Before trial 1, stock A is randomly assigned a state. With the price update in trial $t > 1$, the state of stock i in this trial remains the same as in trial t - 1 with probability 0.8, but switches with probability 0.2. More formally and in line with Frydman et al. (2014), if $s_{i,t} \in \{good, bad\}$ is the state of the i-th stock in trial t, then the state switches as follows:

	$s_{i,t}$ = *good*	$s_{i,t}$ = *bad*
$s_{i,t-1}$ = *good*	0.8	0.2
$s_{i,t-1}$ = *bad*	0.2	0.8

It is expected that participants will infer the states of the stocks from the observed price paths. The states of the stocks are never revealed to them. The same set of realized states and prices is used for all participants, to facilitate comparability. The optimal trading strategy for a risk-neutral Bayesian investor whose objective is to maximize the expected value of her take-home earnings is formalized in Frydman et al. (2014).

Financial incentives: At the end of each of the four sessions in the test and retest parts of the study, participants' holdings of the stock A are liquidated, and the cash value of their position is recorded. Participants' incentives depend on the final value of their portfolio at the end of each session. Specifically, if the total value of a participant's cash and risky asset holdings at the end of session 1 is X_1, the total value of her cash and risky asset holdings at the end of session 2 is X_2, at the end of session 3 is X_3, and at the end of session 4 is X_4 (in experimental currency), then her take-home pay in actual dollars is $10 + (X_1 + X_2 + X_3 + X_4)/48$. The structure of the financial incentives was communicated to the participants before the start of the experiment (see Appendix C).

1.5. ToM-Based Intervention Using the Nencki Affective Picture System (NAPS)

In the second trading session of both the test and retest treatments, participants were asked to recommend purchases or (short-) sales to a client. Participants were shown photographs of 21 clients (see Figure 2), among which they could select one whom they would advise.

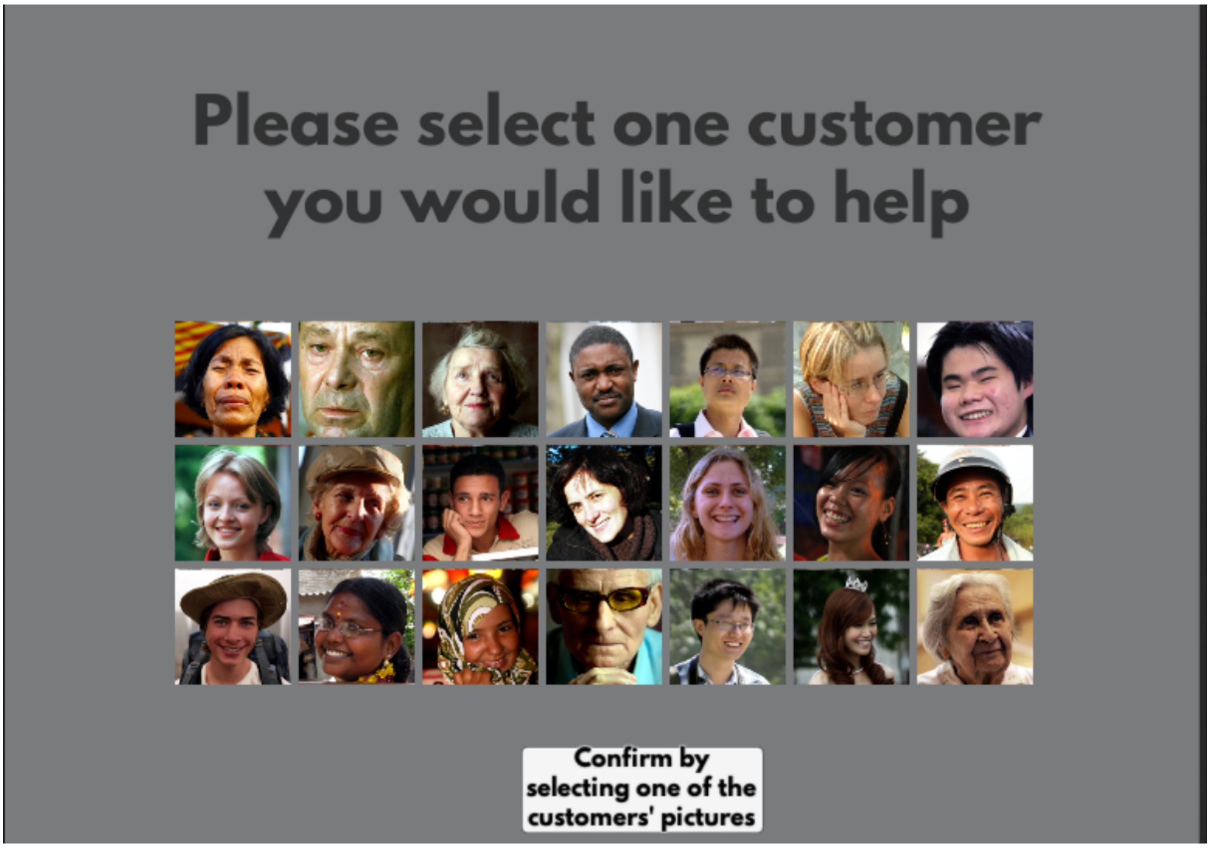


Figure 2. Advisee selection screen.

The photographs were selected from the Nencki Affective Picture System (NAPS, Marchewka et al., 2014).[3] The photographs were chosen to ensure diversity, as to age, gender and ethnic background. Sufficiently high levels of valence, arousal, and approach/avoidance, as reported in NAPS Ratings (ibid.), were additional inclusion criteria for the images used in this study.[4].

[3] The approval to use NAPS images was obtained from the Laboratory of Brain Imaging (LOBI), Neurobiology Center, The Nencki Institute of Experimental Biology on 2.15.2017.

[4] The measures of valence, approach/avoidance, and arousal were obtained based on the reports of 204 healthy volunteers whose demographic characteristics (119 women, 85 men; mean age = 23.9 years, SD = 3.4) and were similar to those of the participants of the current study. The ratings were measured using three 9-point Likert scales. On the valence scale, participants were asked to complete the sentence, "You are judging this image as ..." (from 1 = 'very negative' to 9 = 'very positive', with 5 = 'neutral'). Next, participants judged motivational direction by completing the sentence, "My reaction to this image is ..." (from 1 = 'to avoid' to 9 = 'to approach', with 5 = 'neutral'). Finally, participants judged the degree of arousal elicited by pictures with the introductory sentence, "Confronted with this image, you are feeling: …" (from 1 = 'relaxed' to 9 = 'aroused', with 5 = 'neutral/ambivalent'). The mean ratings of the 21 photographs retained for our training sessions are as follows: valence (M = 6.23, SD = 1.31), approach/avoidance (M = 5.96, SD = 1.24), arousal (M = 4.62, SD = 1.46).

After selecting a client (an advisee), participants trade on behalf of their advisee using the trading dashboard shown in Figure 3. One distinct characteristic of this trading session is that the advisee never holds on to investments for more than one trial. For this reason, the disposition effect in this trading session is not calculated; the trading session functions merely as part of the ToM-based experimental intervention.

In the subsequent trading session, participants may trade both for themselves and on behalf of the advisee (see Figure 4). Participants can choose trades for themselves ("Your A") *separately from those* for the advisee ("Other's A"). Market fluctuations for stocks "Your A" and "Other's A" are, however, perfectly correlated. As shown in Figure 4, the (potentially different) cash positions of the participant as well as of the advisee are displayed separately. In the acquisition price cells, the prices are displayed at which the stock was purchased for oneself (the advisor) and/or for the chosen "Other" (the advisee). But since the advisee is never invested for more than one trial, the acquisition price for the advisee is always reset to "Not Available" – indicated with a dash. At the end of the session, the disposition effect is calculated and reported only for the trades placed by the advisor.

A fourth session follows, which repeats the first session, where participants trade for themselves only (see Figure 1).

The same trading sessions are administered in the retest treatment four weeks after the original test experiment.

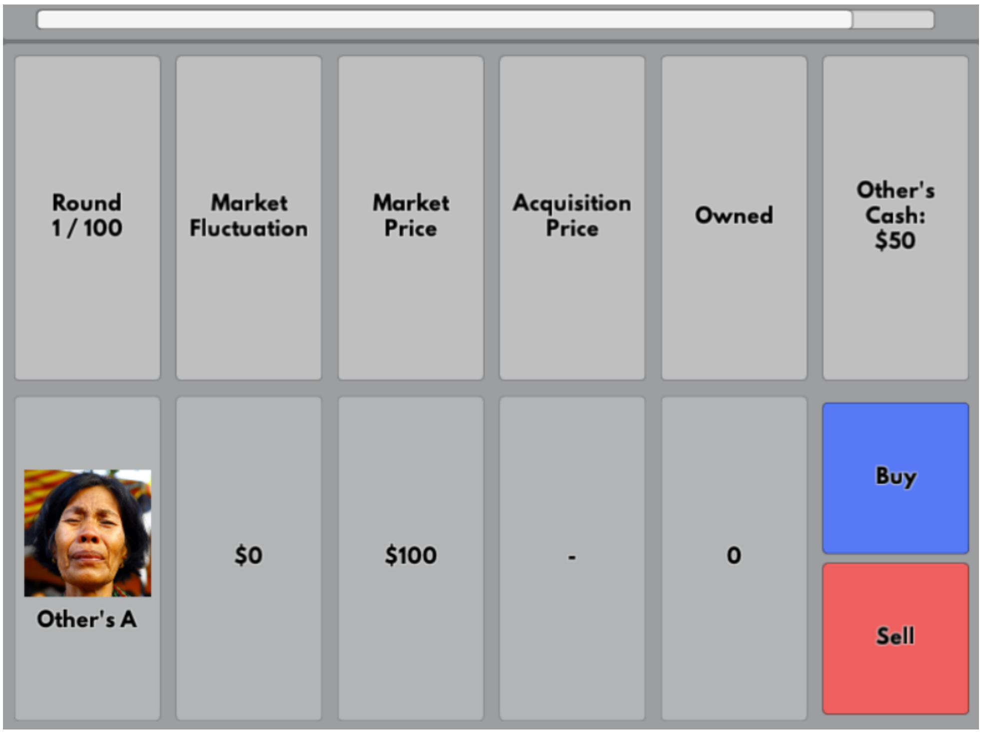


Figure 3. Trading dashboard in Session 2.
Here, participants trade on behalf of their advisee ("Other").

1.6. Clinical Tests of Social Cognition

We administered the Awareness of Social Inference Tests - Revised (TASIT-R) (Flanagan et al., 2011). Details of content and purpose of the test are provided in the Appendix D. Importantly, the subscales of the TASIT-R test allow us to separate the dimensions of ToM associated with emotional evaluation and social cognition.

Test subscales are:

Figure 4. Trading dashboard in Session 3

Participants trade on their own as well as for their advisee's account; the advisee immediately offsets her position after one round, and hence, any investment in the security has to be renewed every round.

1. The Emotion Evaluation Test (EET);
2. The Social Inference–Minimal (SI-M) test; and
3. The Social Inference–Enriched (SI-E) test.

EET focuses on affective side of human cognition (emotions and empathy), while SI-M and SI-E focus on cognitive aspects of social interaction. The results for each of the three TASIT-R tests relate to distinct components of ToM. All three subscales in the TASIT-R test have test and retest versions (referred to as 'form A' and 'form B'), which allow us to calibrate the test and retest treatments in our study.

1.7. Demographic Questionnaire

An online demographic questionnaire is used to record the age, gender of participants, and their trading experience.

1.8. Psychophysiological Apparatus

Eye movements were recorded using a table-mounted eye tracking system (Tobii TX300) with a temporal resolution of 300 Hertz and a screen resolution of 1920 x 1080 pixels (see Appendix A for more complete technical details associated with the eye tracking technology used). At the average viewing distance of 65 cm from the screen (range: 50-80 cm), binocular accuracy of the eye tracking system was 0.4 degrees and precision was 0.14 degrees[5]. Eye fixations were computed using the velocity-based I-VT algorithm (Komogortsev et al., 2010). For each trial, the eye tracking measures we calculated in relation to the selected as areas of interest (AOIs) displaying the acquisition price and the overall trading dashboard (see Appendices A and B).

The experiment was conducted in light-controlled dimly lit sound-proof booths. Participants sat on height-adjustable chairs with their head supported by a height-adjustable ophthalmological chin rest. At the beginning of the experiment, the eye tracker was calibrated using a nine-point fixation technique, which is the most rigorous calibration technique for the device used. This calibration adjusts for participants' individual differences in eye characteristics and participants' seating position.

1.9. Miscellaneous Protocol Details

Ethics approval was obtained from Monash University, where the research study was conducted[6]. An explanatory statement was presented before administering the main experimental task. Participants read the explanatory statement and signed the consent form, thereby approving the use of the de-identified data collected in this study.

The experiment was administered in the experimental laboratory (Monash Business Behavioural Laboratory) of the Monash Business School. Test and retest treatments of the study

[5] For more detailed product specifications, please refer to Appendix A and the product description on the manufacturer's website: http://www.tobiipro.com/product-listing/tobii-pro-tx300/

[6] Ethics Number CF16/346 - 2016000160

followed the same protocol. Upon signing the consent form, participants accessed the online stock trading task using a web browser link administered via Tobii Studio (version 3.4.5) software on the Tobii TX300 eye tracking systems. A maximum of six participants were trained and tested simultaneously, in separate booths with Tobii TX300 eye tracking systems. Thus, participants could not observe what other participants decided.

The remainder of the protocol was as follows. First, participants read the instructions of the stock trading game (see Appendix C) supplemented with relevant screenshots and a concise explanation of the trading dashboard. While familiarizing themselves with the rules of the game, participants could ask clarifying questions. Next, participants were administered an introductory trading session, in which they were given thirty 10-second experimental trials where they traded for themselves. Feedback in the form of the disposition effect score and the amount (in experimental currency) earned was provided on the computer screen at the end of this introductory session. This was the only session in which the participants received a feedback in terms of a disposition effect score. The introductory trading session took approximately five minutes to complete. This was then followed by the four main sessions in the test sitting.

The first session tested participant's susceptibility to the disposition effect, as per the description provided in the Materials and Apparatus section (see Figure 1). In the second trading session, the participant (the advisor) chose an advisee (Figure 2) and traded on her behalf (Figure 3). The advisor is awarded a fraction of the earnings (25%) of the advisee. In the third trading session, the advisor trades both for his own account and on behalf of the advisee (Figure 4). As in Session 2, the advisee never holds on to investments for more than one trial. That is, gains and losses are realized immediately, thus not allowing the disposition effect to actualize. Compensation is based on the final value of one’s own account and a percentage of the earnings of the advisee. In the fourth session, the first session is repeated and participants trade for themselves only.

Each trading session took approximately 17 minutes to complete. Apart from their cash position, the participants did not receive any other form of feedback after the completion of a trading session.

Finally, participants were administered three subscales of the TASIT-R test (form A) in the following order: EET, SI-M, and SI-E. The completion of all three TASIT-R tests took approximately 45 minutes. Participants were also asked to complete a short demographic questionnaire, which took approximately one minute of their time.

Four weeks later, the retest treatment of the study was administered, which consisted of the same four trading sessions, in the same order, followed by the alternate versions (form B) of the three TASIT-R subscales: EET, SI-M, and SI-E.

Participants signed two separate consent forms for the test and retest treatments. Taking into consideration the time it took the participants to familiarize with the instructions of the trading task, test and retest sittings each lasted for up to 75 minutes[7].

2. Detailed Results

2.1. Descriptive Statistics.

The descriptive statistics for the level of the disposition effect and the scores on TASIT-R subscales (forms A and B) are reported in Table 1.

[7] This does not include the time for completion of the three subscales of the TASIT-R test. On average, the test (retest) sitting, including the three subscales of the TASIT-R test took 120 minutes to complete.

Table 1. Descriptive statistics
(N = 68)

Session	Disposition Effect Mean	Disposition Effect Std. Deviation	TASIT-R subscales	Mean	Std. Deviation
S_1	0.12	0.27	EET(A)	23.04	3.23
So_1	0.12	0.28	SI-M(A)	46.12	7.19
S_2	0.11	0.28	SI-E(A)	47.19	8.08
S_3	0.16	0.25	EET(B)	22.46	2.46
So_2	0.12	0.30	SI-M(B)	46.94	6.47
S_4	0.08	0.30	SI-E(B)	47.66	6.96

Note: See text for meaning of Session and TASIT-R subscale identifiers

Here, sessions are denoted as follows:

- S_1 is the first trading session, in which participants trade only for themselves;
- O_1 is the second trading session, in which participants trade only for the advisee, who always realizes gains (or losses) immediately after the end of an investment trial (as such, no disposition effect is recorded during this session);
- So_1 is the third trading session, in which participants trade both for themselves and for the advisee;
- S_2 is the fourth trading session, which is identical to S_1, where participants trade only for themselves;
- S_3 is the fifth trading session, and the first trading session of the retest treatment of the study; this session is undertaken four weeks after the completion of the test treatment of the study and is identical to the session S_1;
- O_2 is the sixth trading session, and the second trading session of the retest treatment (identical to O_1);

- *So_2* is the seventh trading session, and the third trading session of the retest treatment (identical to So1); and
- *S_4* is the eighth trading session, and the fourth and last trading session of the retest treatment (identical to S_2).

Ordinary Least Square (OLS) regression analysis is used to test the two main predictions in the study. These are:

1. The scores on clinical tests that focus on affective components of social cognition (compared to purely cognitive components) have less influence on training performance (*Hypothesis 1*);
2. Participants with higher ToM scores will be more susceptible to the ToM-based cognitive training and thereby will be able to more substantially reduce their level of disposition effect (*Hypothesis 2*).

Recall that complete behavioral data are available for 68 participants (N = 68).

To justify the use of OLS, a number of diagnostic checks were performed: histogram and probability plots of standardized residuals, as well as scatterplots of standardized residuals against standardized predicted values, were used to ascertain that the assumptions of normality, linearity, and homoscedasticity of residuals were not violated.[8]

2.2. Predicting the level of disposition effect across trading sessions

OLS regression analysis is used to discover association between individual scores on one of the three TASIT-R subscales, i.e. EET, SI-M, or SI-E, and the level of disposition effect in a session or, alternatively, the change in the level of disposition effect across two sessions. A different version (form A or B) of the TASIT-R subscales is used, depending on whether the session pertained to the

[8] The results of these tests can be provided upon request.

test or retest treatment. In the analyses below, form A TASIT-R subscales (pertaining to the test treatment) are coded as EET(A), SI-M(A), and SI-E(A); form B subscales (for the retest treatment) are coded as EET(B), SI-M(B), and SI-E(B).

Results of the OLS regression analyses for the level of the disposition effect, per session, including non-standardized (*B*) and standardized (β) regression coefficients for all predictors (the scores on EET(A), EET(B), SI-M(A), SI-M(B), SI-E(A) and SI-E(B) tests), as well as coefficients of determination (R^2), are reported in Table 2.

Table 2. Predictors of disposition effect level across trading sessions

Regression coefficients (unstandardized and standardized), standard errors, t-statistics, p-values and coefficients of determination for OLS analyses using *EET, SI-M*, and *SI-E* to predict the disposition effect across sessions ($N = 68$).

Session	Predictor variables	***B***	***SE***	***β***	***t***	***p***	***R²***	***Lower CI***	***Upper CI***
S_1									
	EET(A)	**0.02**	**0.01**	**0.22**	**1.83**	**0.07**	**0.05**	**-0.002**	**0.04**
	EET(B)	0.001	0.01	0.01	0.09	0.93	0.000	-0.03	0.03
	SI-M(A)	0.01	0.01	0.14	1.17	0.25	0.02	-0.004	0.02
	SI-M(B)	0.01	0.01	0.19	1.58	0.12	0.04	-0.002	0.02
	SI-E(A)	0.01	0.004	0.14	1.11	0.27	0.02	-0.004	0.01
	SI-E(B)	**0.01**	**0.01**	**0.25**	**2.08**	**0.04**	**0.06**	**0.000**	**0.02**
So_1									
	EET(A)	-0.01	0.01	-0.08	-0.68	0.50	0.01	-0.03	0.01
	EET(B)	0.000	0.01	-0.002	-0.01	0.99	0.000	-0.03	0.03
	SI-M(A)	-0.002	0.01	-0.05	-0.37	0.71	0.002	-0.01	0.01
	SI-M(B)	-0.000	0.01	-0.001	-0.01	0.99	0.000	-0.01	0.01
	SI-E(A)	-0.001	0.004	-0.03	-0.27	0.79	0.001	-0.01	0.01
	SI-E(B)	0.000	0.01	-0.003	-0.03	0.98	0.000	-0.01	0.01
S_2									
	EET(A)	0.01	0.01	0.07	0.60	0.55	0.01	-0.02	0.03
	EET(B)	0.01	0.01	0.11	0.86	0.40	0.01	-0.02	0.04
	SI-M(A)	0.001	0.01	0.02	0.13	0.90	0.000	-0.01	0.01
	SI-M(B)	-0.003	0.01	-0.06	-0.47	0.64	0.003	-0.01	0.01
	SI-E(A)	-0.002	0.004	-0.05	-0.42	0.68	0.003	-0.01	0.01
	SI-E(B)	-0.001	0.01	-0.02	-0.14	0.89	0.000	-0.01	0.01
S_3									
	EET(A)	0.01	0.01	0.13	1.03	0.31	0.02	-0.01	0.03
	EET(B)	0.002	0.01	0.02	0.19	0.85	0.001	-0.02	0.03
	SI-M(A)	**0.01**	**0.004**	**0.28**	**2.33**	**0.02**	**0.08**	**0.001**	**0.02**
	SI-M(B)	**0.01**	**0.004**	**0.29**	**2.48**	**0.02**	**0.09**	**0.002**	**0.02**
	SI-E(A)	**0.01**	**0.004**	**0.30**	**2.52**	**0.01**	**0.09**	**0.002**	**0.02**
	SI-E(B)	**0.01**	**0.004**	**0.28**	**2.37**	**0.02**	**0.08**	**0.002**	**0.02**
So_2									
	EET(A)	0.000	0.01	0.003	0.02	0.98	0.000	-0.02	0.02
	EET(B)	-0.002	0.02	-0.02	-0.15	0.88	0.000	-0.03	0.03
	SI-M(A)	0.01	0.01	0.20	1.66	0.10	0.04	-0.002	0.02
	SI-M(B)	0.01	0.01	0.13	1.08	0.29	0.02	-0.01	0.02
	SI-E(A)	0.01	0.004	0.17	1.42	0.16	0.03	-0.003	0.02
	SI-E(B)	0.01	0.01	0.13	1.05	0.30	0.02	-0.01	0.02
S_4									
	EET(A)	0.01	0.01	0.05	0.44	0.66	0.003	-0.02	0.03
	EET(B)	0.01	0.02	0.04	0.31	0.76	0.001	-0.03	0.03
	SI-M(A)	-0.01	0.01	-0.16	-1.35	0.18	0.03	-0.02	0.003
	SI-M(B)	**-0.01**	**0.01**	**-0.23**	**-1.91**	**0.06**	**0.05**	**-0.02**	**0.000**
	SI-E(A)	**-0.01**	**0.004**	**-0.23**	**-1.94**	**0.06**	**0.05**	**-0.02**	**0.000**
	SI-E(B)	**-0.01**	**0.01**	**-0.25**	**-2.08**	**0.04**	**0.06**	**-0.02**	**0.000**

Here we summarize the results only for those tests which attained significance at the 10% level.

In the first trading session ($S_{1,}$), when participants make investment decisions for themselves, EET(A) accounts for a significant 5% variance in the disposition effect; $F(1,66) = 3.35$, $p = 0.07$. Participants' disposition effect increases by 0.02 (see non-standardized *B* coefficient in Table 2) for each point of the total EET(A) score. Similarly, SI-E(B) is found to be significantly and positively associated with the observed level of disposition effect, accounting for 6% of its variance; $F(1,66) = 4.32$, $p = 0.04$.

In the fifth trading session ($S_{3,}$, the first trading session of the retest sitting), when participants make investment decisions for themselves, SI-M(A) accounts for a significant 8% variance in the disposition effect; $F(1,66) = 5.45$, $p = 0.02$. Likewise, SI M(B) accounts for 9% variance ($F(1,66) = 6.17$, $p = 0.02$); SI-E(A) for 9% variance ($F(1,66) = 6.33$, $p = 0.01$); SI-E(B) for 8% variance ($F(1,66) = 5.61$, $p = 0.02$).

In the eighth trading session (S_4, the fourth and last trading session of the retest sitting), three TASIT-R subscales, i.e. SI-M(B), SI-E(A) and SI=E(B), are found to be significantly and negatively associated with the level of disposition effect observed in S_4: SI-M(B) explains 5% variance of the disposition effect ($F(1,66) = 3.65$, $p = 0.06$); SI-E(A) for 5% ($F(1,66) = 3.77$, $p = 0.06$); SI-E(B) for 6% variance ($F(1,66) = 4.32$, $p = 0.04$).

2.2. Predicting the Change in Disposition Effect across Trading Sessions

We next investigate the relationship between the scores of three subscales of the TASIT-R tests and the **changes** in the disposition effect between trading sessions. The latter are calculated as negative post-pre values, which implies that a *positive value reflects a reduction in the disposition effect*. Non-standardized (*B*) and standardized (β) regression coefficients and coefficients of determination (R^2) are reported in Table 3 and visualized schematically in Figure 6.

Table 3. Reduction in disposition effect across sessions

Explanatory variables are calculated as post-pre disposition effect values across two trading sessions. Reported are regression coefficients (unstandardized and standardized), standard errors, t-statistics, p-values and coefficients of determination for the OLS analyses using *EET*, *SI-M*, and *SI-E* (tests and retests) as predictors *(N = 68)*.

Post-Pre Difference in DE	Predictor variables	*B*	*SE*	*ß*	*t*	*p*	R^2	*Lower CI*	*Upper CI*
So_1-S_1									
	EET(A)	**-0.03**	**0.01**	**-0.28**	**-2.33**	**0.02**	**0.08**	**-0.05**	**-0.004**
	EET(B)	-0.001	0.02	-0.01	-0.10	0.92	0.000	-0.03	0.03
	SI-M(A)	-0.01	0.01	-0.17	-1.41	0.16	0.03	-0.02	0.003
	SI-M(B)	-0.01	0.01	-0.17	-1.43	0.16	0.03	-0.02	0.003
	SI-E(A)	-0.01	0.01	-0.15	-1.26	0.21	0.02	-0.02	0.003
	SI-E(B)	**-0.01**	**0.01**	**-0.23**	**-1.90**	**0.06**	**0.05**	**-0.02**	**0.001**
S_2- So_1									
	EET(A)	0.01	0.01	0.14	1.14	0.26	0.02	-0.01	0.04
	EET(B)	0.01	0.02	0.10	0.78	0.44	0.01	-0.02	0.04
	SI-M(A)	0.002	0.01	0.05	0.44	0.66	0.003	-0.01	0.01
	SI-M(B)	-0.002	0.01	-0.05	-0.41	0.68	0.003	-0.01	0.01
	SI-E(A)	-0.001	0.01	-0.02	-0.13	0.90	0.000	-0.01	0.01
	SI-E(B)	-0.001	0.01	-0.01	-0.11	0.92	0.000	-0.01	0.01
S_2-S_1									
	EET(A)	-0.01	0.01	-0.12	-1.01	0.32	0.02	-0.04	0.01
	EET(B)	0.01	0.02	0.08	0.68	0.50	0.01	-0.02	0.04
	SI-M(A)	-0.01	0.01	-0.11	-0.89	0.38	0.01	-0.02	0.01
	SI-M(B)	**-0.01**	**0.01**	**-0.22**	**-1.80**	**0.08**	**0.05**	**-0.02**	**0.001**
	SI-E(A)	-0.01	0.01	-0.16	-1.34	0.19	0.03	-0.02	0.003
	SI-E(B)	**-0.01**	**0.01**	**-0.23**	**-1.91**	**0.06**	**0.05**	**-0.02**	**0.000**
S_3-S_1									
	EET(A)	-0.01	0.01	-0.11	-0.90	0.37	0.01	-0.03	0.01
	EET(B)	0.001	0.01	0.01	0.08	0.94	0.000	-0.03	0.03
	SI-M(A)	0.004	0.004	0.11	0.91	0.37	0.01	-0.01	0.01
	SI-M(B)	0.003	0.01	0.08	0.62	0.54	0.01	-0.01	0.01
	SI-E(A)	0.004	0.004	0.14	1.13	0.26	0.02	-0.003	0.01
	SI-E(B)	0.000	0.01	0.01	0.05	0.96	0.000	-0.01	0.01
S_3-S_2									
	EET(A)	0.003	0.01	0.03	0.26	0.79	0.001	-0.02	0.03
	EET(B)	-0.01	0.02	-0.08	-0.63	0.53	0.01	-0.04	0.02
	SI-M(A)	**0.01**	**0.01**	**0.20**	**1.69**	**0.10**	**0.04**	**-0.002**	**0.02**
	SI-M(B)	**0.01**	**0.01**	**0.28**	**2.39**	**0.02**	**0.08**	**0.002**	**0.03**
	SI-E(A)	**0.01**	**0.01**	**0.28**	**2.37**	**0.02**	**0.08**	**0.002**	**0.02**
	SI-E(B)	**0.01**	**0.01**	**0.24**	**1.98**	**0.05**	**0.06**	**0.000**	**0.02**
So_2-S_3									
	EET(A)	-0.01	0.01	-0.09	-0.75	0.45	0.01	-0.03	0.02
	EET(B)	-0.01	0.02	-0.04	-0.28	0.78	0.001	-0.04	0.03
	SI-M(A)	-0.01	0.01	-0.03	-0.21	0.84	0.001	-0.01	0.01
	SI-M(B)	-0.01	0.01	-0.10	-0.83	0.41	0.01	-0.02	0.01
	SI-E(A)	-0.003	0.01	-0.07	-0.54	0.59	0.004	-0.01	0.01
	SI-E(B)	-0.004	0.01	-0.09	-0.77	0.45	0.01	-0.02	0.01
S_4- So_2									
	EET(A)	0.01	0.01	0.05	0.38	0.71	0.002	-0.02	0.03
	EET(B)	0.01	0.02	0.05	0.42	0.68	0.003	-0.03	0.04
	SI-M(A)	**-0.02**	**0.01**	**-0.33**	**-2.81**	**0.01**	**0.11**	**-0.03**	**-0.004**
	SI-M(B)	**-0.02**	**0.01**	**-0.32**	**-2.76**	**0.01**	**0.10**	**-0.03**	**-0.01**
	SI-E(A)	**-0.02**	**0.01**	**-3.62**	**-3.16**	**0.002**	**0.13**	**-0.02**	**-0.01**
	SI-E(B)	**-0.02**	**0.01**	**-0.34**	**-2.90**	**0.01**	**0.11**	**-0.03**	**-0.01**
S_4-S_3									
	EET(A)	-0.01	0.01	-0.04	-0.36	0.72	0.002	-0.03	0.02
	EET(B)	0.002	0.02	0.02	0.14	0.89	0.000	-0.03	0.04
	SI-M(A)	**-0.02**	**0.01**	**-0.34**	**-2.97**	**0.004**	**0.12**	**-0.03**	**-0.01**
	SI-M(B)	**-0.02**	**0.01**	**-0.41**	**-3.66**	**0.001**	**0.17**	**-0.03**	**-0.01**
	SI-E(A)	**-0.02**	**0.01**	**-0.42**	**-3.72**	**0.001**	**0.17**	**-0.03**	**-0.01**
	SI-E(B)	**-0.02**	**0.01**	**-0.42**	**-3.74**	**0.001**	**0.18**	**-0.03**	**-0.01**
S_4-S_1									
	EET(A)	-0.01	0.01	-0.12	-1.00	0.32	0.02	-0.04	0.01
	EET(B)	0.003	0.02	0.02	0.18	0.86	0.001	-0.03	0.04
	SI-M(A)	**-0.01**	**0.01**	**-0.24**	**-2.03**	**0.05**	**0.06**	**-0.02**	**0.000**
	SI-M(B)	**-0.02**	**0.01**	**-0.33**	**-2.86**	**0.01**	**0.11**	**-0.03**	**-0.01**
	SI-E(A)	**-0.01**	**0.01**	**-0.29**	**-2.49**	**0.02**	**0.09**	**-0.02**	**-0.003**
	SI-E(B)	**-0.02**	**0.01**	**-0.39**	**-3.46**	**0.001**	**0.15**	**-0.03**	**-0.01**

Here, we mention only the significant (at 10% level) findings.

When testing the reduction in the disposition effect *between the first trading session* (trading for oneself, S_1) *and the third trading session* (trading both for oneself and others, So_1), i.e. $(So_1\text{-}S_1)$, EET(A) and SI-E(B) are found to account for a significant *reduction* of the disposition effect explaining 8% and 5% of the variation in the change ($F(1{,}66) = 5.43$, $p = 0.02$ and $F(1{,}66) = 3.59$, $p = 0.06$, respectively).

Next, SI-M(B) and SI-E(B) are found to explain the *reduction* in the disposition effect *between the first and the fourth trading sessions*, i.e. $(S_2\text{-}S_1)$, on marginally significant levels with 5% and 5% of variance explained ($F(1{,}66) = 3.22$, $p = 0.08$ and $F(1{,}66) = 3.65$, $p = 0.06$).

The analysis of the change in disposition effect before and after the four-week test-retest interval shows a pronounced treatment washout. Thus, between *the fourth trading session and the fifth trading session* (i.e. $(S_3\text{-}S_2)$), SI-M(A), SI-M(B), SI-E(A) and SI-E(B) are found to be significantly associated with an *increase* in the disposition effect. SI-M(A) explains 4% of the variance in the increase of the disposition effect ($F(1{,}66) = 2.85$, $p = 0.096$). SI-M(B) explains 8% of the variation in the change ($F(1{,}66) = 5.73$, $p = 0.02$); SI-E(A) explains 8% of the variance ($F(1{,}66) = 5.59$, $p = 0.02$); SI-E(B) explains 6% of the variance ($F(1{,}66) = 3.91$, $p = 0.05$).

Between *the seventh trading session* and *the eighth trading session* (i.e. $(S_4\text{- }So_2)$), SI-M(A), SI-M(B), SI-E(A), and SI-E(B) all account for a significant *reduction* in the disposition effect, with 11%, 10%, 13% and 11% of the variance explained, respectively. The *F*-statistics and significance level(s) are $F(1{,}66) = 7.87$, $p = 0.01$ for SI-M(A); $F(1{,}66) = 7.63$, $p = 0.01$ for SI-M(B); $F(1{,}66) = 9.98$, $p = 0.002$ for SI-E(A); $F(1{,}66) = 8.43$, $p = 0.01$ for SI-E(B).

Between *the fifth trading session* and *the eighth trading session*, both in the retest treatment ((i.e. $(S_4\text{- }S_3)$), SI-M(A), SI-M(B), SI-E(A), and SI-E(B) again are found to explain a significant fraction of the reduction in the disposition effect, with 12%, 17%, 17%, and 18% of the variance explained. The *F*-statistics and significance level(s) for SI-M(A) are $F(1{,}66) = 8.79$, $p = 0.004$; for

SI-M(B) are $F(1,66) = 13.42, p < 0.001$; for SI-E(A) are $F(1,66) = 13.86, p < 0.001$; and for SI-E(B) are $F(1,66) = 14.01, p < 0.001$.

When predicting the improvement in the disposition effect between *the first trading session* (playing for oneself) and *the eighth trading session* (the last trading session, playing for oneself) (i.e. (S_4- S_1)), SI-M(A) accounted for a significant 6 % variance in the change of disposition effect, $F(1,66) = 4.13$, $p = 0.05$. SI-M(B) accounted for a significant 11 % variance in the change of disposition effect, $F(1,66) = 8.21$, $p = 0.01$. SI-E(A) accounted for a significant 9% variance in the change of disposition effect, $F(1,66) = 6.22$, $p = 0.02$. SI-E(B) accounted for a significant 15% variance in the change of disposition effect, $F(1,66) = 11.93, p < 0.001$.

3.3 Testing for the reduction in disposition effect across high- and low-ToM Groups

The results reported before demonstrate that individuals with higher social inference skills measured using SI-M and SI-E subscales of the TASIT-R test are more responsive to the ToM-based intervention, while higher scores on the emotional evaluation subscale (EET) do not lead to a significant cognitive training effect. This supports the *Hypothesis 1* of this study. As shown in Table 2 and Figure 5, among the two TASIT-R social inference subscales, only SI-E consistently (which means: based on both the form A and form B scores) predicted the reduction in the disposition effect upon the second intervention.

To provide a more nuanced understanding of the changes in the disposition effect across low- and high-ToM groups, we averaged the test-retest scores across the form A and B SI-E tests, after which we did a median split (Lacobucci et al., 2015). The mean levels of disposition effect before and after interventions in the low- and high-TOM groups are shown in Figure 7.

A series of paired-samples *t*-tests are conducted to test for changes in disposition effect scores pre- and post-intervention in the resulting low- and high-ToM groups. We verified that the assumption of homogeneity of variance is met for disposition effect scores across all sessions depicted in Figure 7, and for both groups.

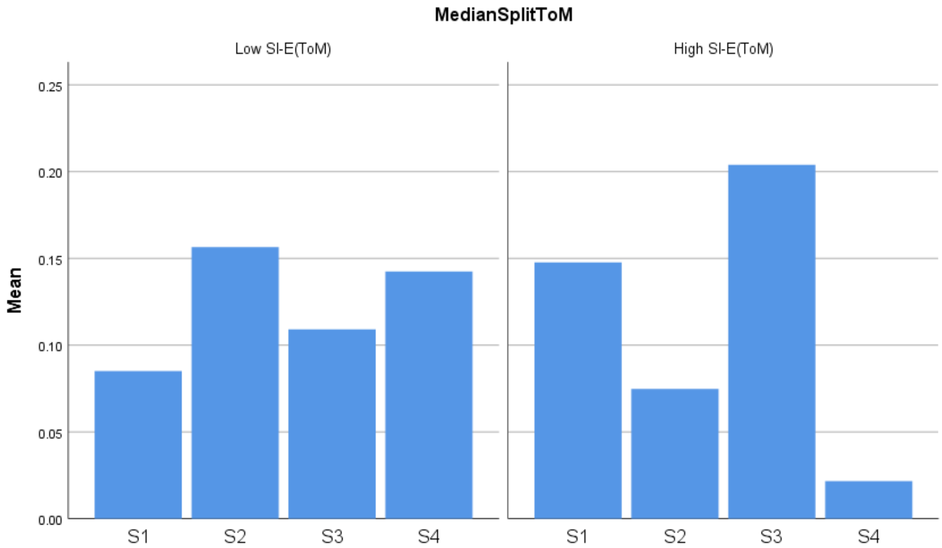


Figure 7. Mean disposition effect levels in 'trading for oneself' sessions stratified by ToM scores

ToM groups are formed based on the median split of the mean scores of SI-E subscale (forms A and B) of the TASIT-R test.

No significant differences in one-tailed paired-samples *t*-tests of the mean disposition effect scores emerge in the *low-ToM* group ($p > 0.10$ for the following tests: S_1 and S_2, $t(32) = -1.78$; S_2 and S_3, $t(32) = 1.12$; S_3 and S_4, $t(32) = -0.82$; S_1 and S_4 $t(32) = -1.26$). Likewise, no significant effect emerges for the *high-ToM* group between sessions S_1 and S_2 ($t(34) = 1.16$, $p = 0.13$). However, between sessions S_2 and S_3 in the high-ToM group, the difference is significant: $t(34) = -2.19$, $p = 0.02$. A *lower* mean disposition score is observed in session S_2 ($M = 0.07$, $SD = 0.30$) which follows the first intervention, than in session S_3 ($M = 0.20$, $SD = 0.27$), the first session after a four-week treatment washout period.

A significant *reduction* in disposition effect is also recorded between sessions S_3 and S_4 ($t(34) = 2.75$, $p = 0.005$). The mean disposition score is lower in session S_4, the last session of the experiment ($M = 0.02$, $SD = 0.35$), than in session S_3, the last session of the test treatment ($M = 0.20$, $SD = 0.27$). The difference of 0.18 amounts to an 89.4% reduction in the disposition effect.

Finally, the comparison of scores obtained in the first (S_1) and last (S_4) trading sessions of the experiment reveal an *85.4% reduction* in the disposition effect; it is significant ($t(35) = 1.79$, $p = 0.04$, one-tailed).

2.4. Isolating the Mechanism Leading to Reduction in Disposition Effect Among Above-Median Social Cognitive Skill Participants

The statistical analyses presented in the previous section have shown a significant reduction of disposition effect for participants with an above-median ToM score, thus supporting the *Hypothesis 2* of the study. The analyses however do not reveal the mechanism behind the observed reduction in the disposition effect. Here, we investigate whether and when successful trainees stop paying attention to the cue that is key to the disposition effect, namely, the purchase price. We measure the selective aspects of participants' attention using eye gaze (Kahneman, 1973; Rehder & Hoffman, 2005): if a participant's eyes are oriented towards an object (e.g., the purchase price), we assume that she is paying more attention to the object than to others on the screen. Following Holmqvist et al. (2011), total fixation count on the trading dashboard is used as the base measure of eye gaze fixations, and fixations on an area of interest (AOI) is counted against this base[9]. Therefore, one AOI is the cell in the trading window displaying the acquisition price (AOI_AP); the other ones are: AOI_DB (entire trading Dashboard), AOI_APS (Own Acquisition Price), AOI_APO (Advisee's Acquisition Price). Our measures capture the degree of attention to the AOI, normalized for total attention[10] paid to all information presented on the trading dashboard.

In Table 4, descriptive statistics on eye gaze results are presented. A number of independent-samples *t*-tests on the eye gazes will now be discussed. These test for significance of differences in several eye gaze ratios across sessions, as well as differences across (high- and low-ToM) groups.

---

[9] We also investigate total fixation durations (total dwell time). This was found to yield the same results (including significance levels) to the ones based on the ratio of fixation counts reported here.

[10] Herein, the reduction of individual attention to the information presented on the trading dashboard may either be associated with individual's disengagement with the task or individual's thought processing which does not necessitate the regular attendance of the information cues presented on the trading dashboard.

Table 4. Eye Gazes.

Descriptive statistics associated with the eye tracking measures ($N = 63$)

Eye tracking measure	Mean (Count or Ratio)	Std. Deviation
S_1_AOI_AP	272.98	115.13
S_1_AOI_DB	2251.75	586.11
S_1_AOI_ratio_APDB	0.13	0.05
So_1_AOI_APO[11]	85.46	54.96
So_1_AOI_APS[12]	200.51	102
So_1_AOI_DB	2302.78	617.11
So_1_AOI_ratio_APODB	0.04	0.03
So_1_AOI_ratio_APSDB	0.09	0.04
S_2_AOI_AP	234.25	111.88
S_2_AOI_DB	1978.60	510.35
S_2_AOI_ratio_APDB	0.12	0.05
S_3_AOI_AP	227.60	119.01
S_3_AOI_DB	2125.90	586.34
S_3_AOI_ratio_APDB	0.11	0.05
So_2_AOI_APO	65.90	47.57
So_2_AOI_APS	153.86	88.35
So_2_AOI_DB	2216.11	586.79
So_2_AOI_APODB	0.03	0.02
So_2_AOI_APSDB	0.07	0.04
S_4_AOI_AP	184.51	125.46
S_4_AOI_DB	1943.29	523.14
S_4_AOI_ratio_APDB	0.10	0.06

Notes: The titles for eye tracking measures consist of (i) *Session* (S_1, So_1,…, S_4), (ii) *Area of Interest* 'AOI', (iii) *level* (e.g. Acquisition Price 'AP', Own Acquisition Price 'APS' (i.e. 'Acquisition Price – Self"), Advisee's Acquisition Price 'APO' (i.e. 'Acquisition Price – Other'), or entire Dashboard 'DB'), or *ratio* (e.g., ratio_APDB stands for 'ratio of AP over DB').

In session S_1, many differences between high- and low-ToM groups are insignificant, such as the ratio of the count of fixations on the acquisition price and on the dashboard (S_1_AOI_ratio_APDB). The lack of significant group differences is also observed in the session

[11] Fixation count on the AOI associated with the acquisition price of the security when trading for others (in the condition when the participants can trade for themselves and others).

[12] Fixation count on the AOI associated with the acquisition price of the security when trading for oneself (in the condition when the participants can trade for themselves and others).

where participants trade both for themselves and others (So_1), for (i) the ratio between count of fixation on (own) acquisition price and on dashboard (So_1_AOI_ratio_APSDB), and (ii) the ratio between count of fixation on the advisee's acquisition price and on dashboard (So_1_AOI_ratio_APODB).

In the last session of the test treatment (S_2), a significant group difference is observed for the ratio between fixation count on the acquisition price and on the dashboard (S_2_AOI_APDB), $t(61) = 2.18$, $p = 0.02$, one-tailed. Figure 8 depicts this in a boxplot.

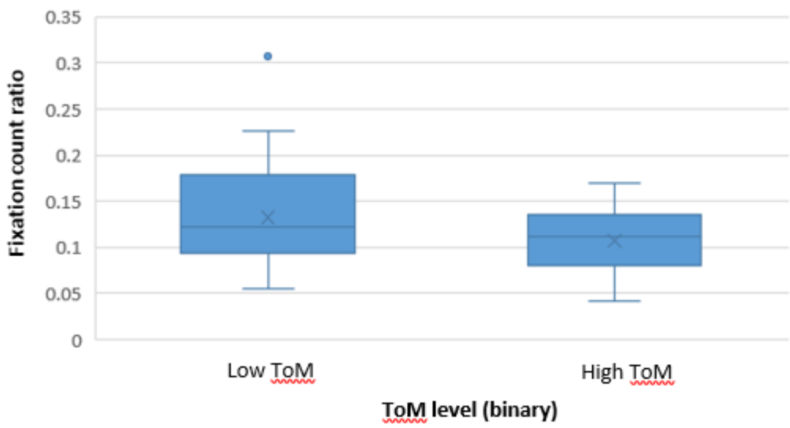


Figure 8. Differences in attention to own acquisition price across ToM levels in the last session of the test treatment (S_2)

Shown are boxplots of the ratios of count of fixations on the acquisition price and on the entire trading dashboard (S_2_ratio_APDB), separately for participants with high- and low-ToM, Session S_2.

In the first experimental session of the retest treatment (coded as Session S_3), no significant group differences are recorded for the ratio between count of fixation on the acquisition price and on the dashboard (S_3_AOI_ratio_APDB). In contrast, in the session of the retest treatment where participants trade both for themselves and others (So_2_AOI_APSDB), highly significant differences between high- and low-ToM groups are found ($t(61) = 3.83$, $p < 0.001$, one-tailed). This is shown graphically with a boxplot in Figure 9. The same happens to the ratio of fixation count on the advisee's acquisition price against total fixation count (So_2_AOI_APODB) ($t(61) = 3.49$, $p < 0.001$, one-tailed).

Figure 10 displays the corresponding boxplots. Finally, in the last trading session of the experiment (S_4_AOI_APDB), highly significant group differences in counts of relative fixation on the acquisition price emerge, $t(61) = 4.94$, $p < 0.001$, one-tailed; see Figure 11.

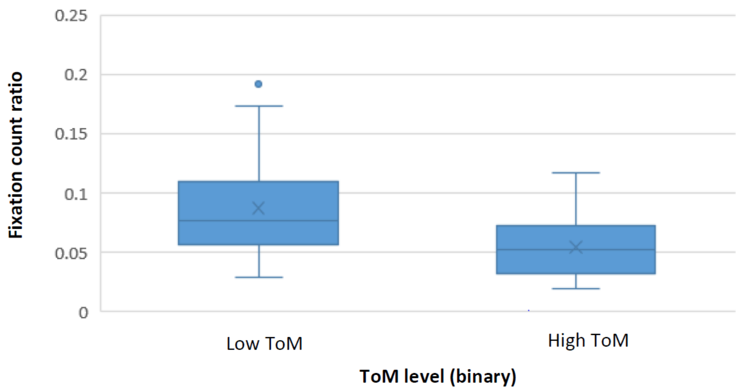


Figure 9. Differences in attention to own acquisition price across ToM levels in the third session after washout period (So_2)
Displayed are boxplots of the ratios of count of fixations on own acquisition price and fixation count on the trading dashboard, stratified by participants' ToM score, Session So_2.

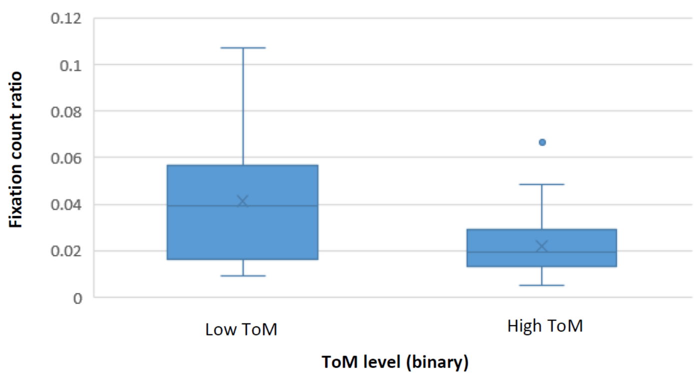


Figure 10. Differences in attention to advisee's acquisition price across ToM levels in the third session after washout period (So_2)
Displayed are boxplots of the ratios of count of fixations on advisee acquisition price against total fixation count, stratified by participants' ToM score, Session So_2.

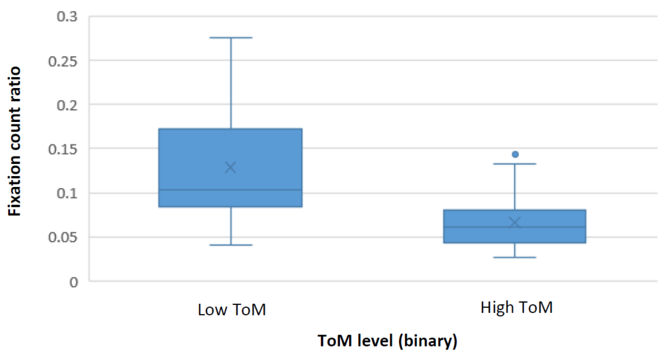


Figure 11. Differences in attention to acquisition price across ToM levels, last session of the experiment (S_4)

Displayed are boxplots of the ratios of count of fixations on the acquisition price and on the trading dashboard, stratified by participants' ToM score, Session S_4.

Next, we report on within-group gaze fixation count results on the acquisition price before and after the interventions in test and retest sittings.

In *low* ToM group, there is a significant *10% increase* in the number of gaze fixations on the acquisition price between post-intervention session (S_4) compared to pre-intervention session (S_3) in the retest sitting ($t(30) = -1.79$, $p = 0.04$). However, there is no significant difference in the number of gaze fixations on the acquisition price between S_1 and S_2, and between S_1 and S_4 in low ToM group.

In *high* ToM group, there is a significant decline in the number of gaze fixations on the acquisition price after the experimental treatment in both test and retest sittings. After the first intervention (S_2-S_1), the number of gaze fixations on the acquisition price decreased by 11.4% ($t(31) = 2.57$, $p = 0.01$); after the second intervention (S_4-S_3), the number of gaze fixations on the acquisition price declined by 36.6% ($t(31) = 6.03$, $p < 0.001$). Overall, from the first to the last experimental sessions (S_4-S_1), the number of gaze fixations on the acquisition price decreased by 45.1% ($t(31) = 6.40$, $p < 0.001$).

Altogether, these findings show that high-ToM participants overall pay less attention to the acquisition price already from the first intervention. This effect is magnified after the four-week washout period, despite a temporary increase in the disposition effect in the first session of the retest treatment. Eye gaze evidently follows the same pattern as the disposition effect score, suggesting that reductions in the disposition effect can be attributed to attention that is turned away from irrelevant information, in this case, the acquisition price.

INTERNET APPENDICES

Appendix A: Technical specifications of the eye tracking measurements[13]

Core parameters	Parameter description	Parameter specifications adopted in the experiment
Measures	Definition of eye tracking measures	Fixation count[14], i.e. the number of times the participant fixates on an AOI, is adopted as the base eye tracking measure. In this study, we used the composite eye tracking measure, which is the ratio of the fixation count on the AOI associated with the purchasing price to the AOI denoting the number of fixations on the trading dashboard.
Apparatus	Sampling procedure	Binocular recording procedure was used (i.e. pupil dilation and eye tracking measures are based on the data acquired from both left and right eyes of the participants)
	Name and produce of the eye tracking device	Tobii TX300, Tobii (Sweden)
	Type of eye tracking device	Desk-mounted
	Sampling rate	300 Hz
	Sampling rate variability	0.3%
	Processing latency	1.0 – 3.3 ms
	Accuracy[15]	0.4^0 – at ideal conditions[16], 0.3^0 - at 25^0 gaze, 0.6^0 - at 30^0 gaze, 0.6^0 – at 1 lux[17], 0.4^0 – at 300 lux, 0.5^0 – at 600 lux, 0.5^0 – at 1000 lux.
	Precision	0.01^0 – with Stamper filter (for more details on the applied Stampe algorithm for noise reduction see Stampe, 1993)
	Eye tracking software used	Tobii Studio 3.4.5
	Chin rest used	Yes
Monitor	Screen size	23‴
	Screen resolution	1920 x 1080 pixel
	Distance between participant and screen	Operating distance: 50-80cm Default distance used in this study: 65cm
Calibration	How many points in calibration	9-point calibration

[13] Informed by Orquin and Holmqvist (2018).

[14] This measure is among the most widely used in eye tracking research (Holmqvist et al., 2011).

[15] The angular average distance from the actual gaze point to the one measured by the eye tracker.

[16] The default experimental setup of this study conforms to the definition of 'accuracy under ideal conditions' outlined in Tobii (2014) as follows: (i) the head movement of the participant is fixed in a chinrest; and (ii) data collected immediately after calibration, in a controlled laboratory environment with constant illumination, with 9 stimuli points (related to the 9-point calibration procedure undertaken in this study) at gaze angle $\leq 18^0$.

.[17] Unit of illuminance and luminous emittance, measuring luminous flux per unit area. One *lux* is equal to one lumen per square metre.

	Amount of recalibration	No recalibration used[18]
Materials	Example image included?	Yes (see Figures 1-4 and Appendix B)
	Participant vision (corrected or not)	corrected-to-normal vision.
Areas of Interest (AOIs)	AOIs used for eye tracking data analysis	The list of AOIs is reported in Table 4 and illustrated in Figures B1, B2, B3 (Appendix B)
Exclusions	Number of trials excluded	None
	Number of participants excluded due to the missing eye tracking data	5 participants (7.35% of the behavioral sample)
	Data quality threshold	A data quality threshold of 85% was used, i.e. at least 85% of the eye tracking data for the practice and for all of the trading sessions had to be present, otherwise the participant was excluded from the sample.
Event detection	What algorithm is used for event detection	The IV-T fixation filter (Komogortsev et al., 2010) was adopted via the selection of global settings in the eye tracking software (Tobii Studio 3.4.5). While the minimal length of fixations did not play a major role in the calculation of processing speed in our study, a rather conservative 60ms threshold was selected within IV-T Tobii filter parameters to define fixations.

[18] As the participants used chin rests, and the duration of the main part of the study (i.e. each experimental treatment) was not excessive (approximately 75 minutes), no recalibration of the eye tracking devices was required.

APPENDIX B: Visual representation of the areas of interest used in the eye tracking analysis

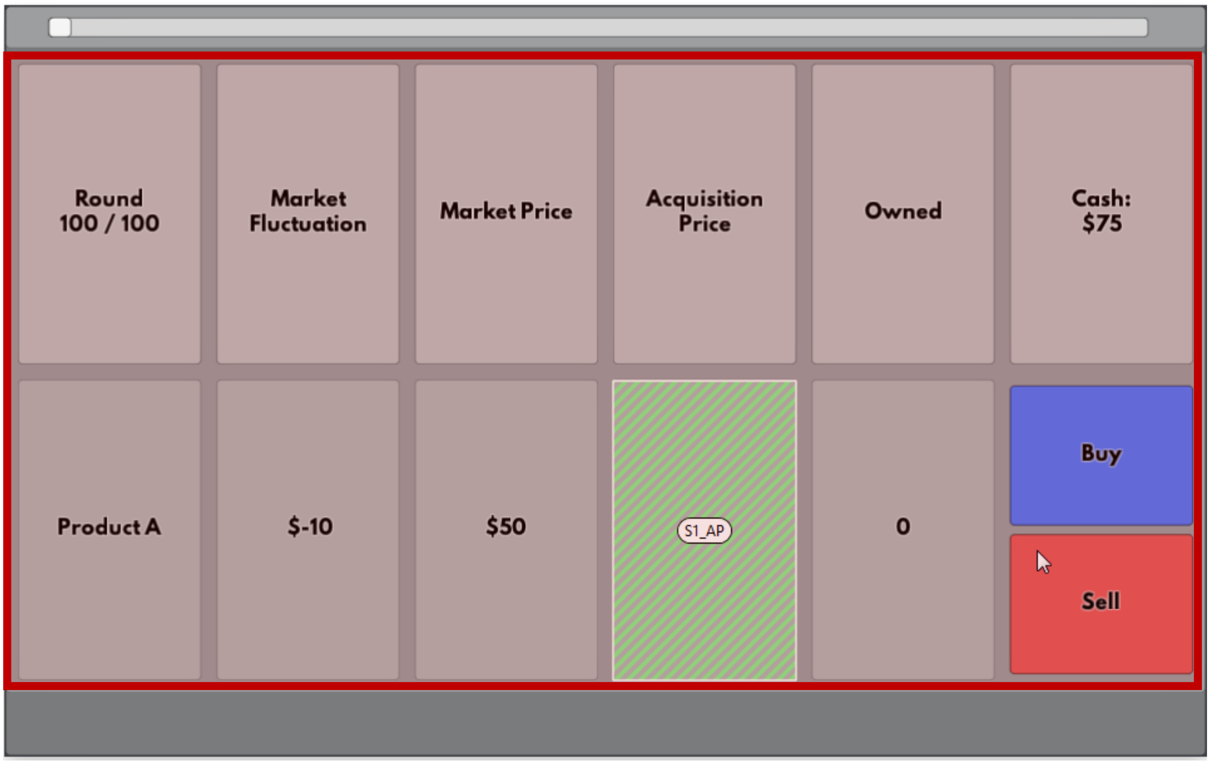


Figure B1. AOIs on the version of the trading dashboard where participants trade for themselves; a smaller AOI covers the area where the acquisition price is indicated, whereas the larger AOI covers the whole dashboard (excluding the progress bar).

Figure B2. AOIs on the version of the trading dashboard where participants trade for others; a smaller AOI covers the area where the acquisition price is indicated, whereas the larger AOI covers the whole dashboard (excluding the progress bar).

Figure B3. AOIs on the version of the trading dashboard where participants trade for both themselves and others; two smaller AOIs cover the areas where the acquisition price of the security purchased for oneself ('Your A') and others ('Other's A') is indicated, whereas the larger AOI covers the whole dashboard (excluding the progress bar).

APPENDIX C: Instructions supplementing the experimental task (Disposition Game)

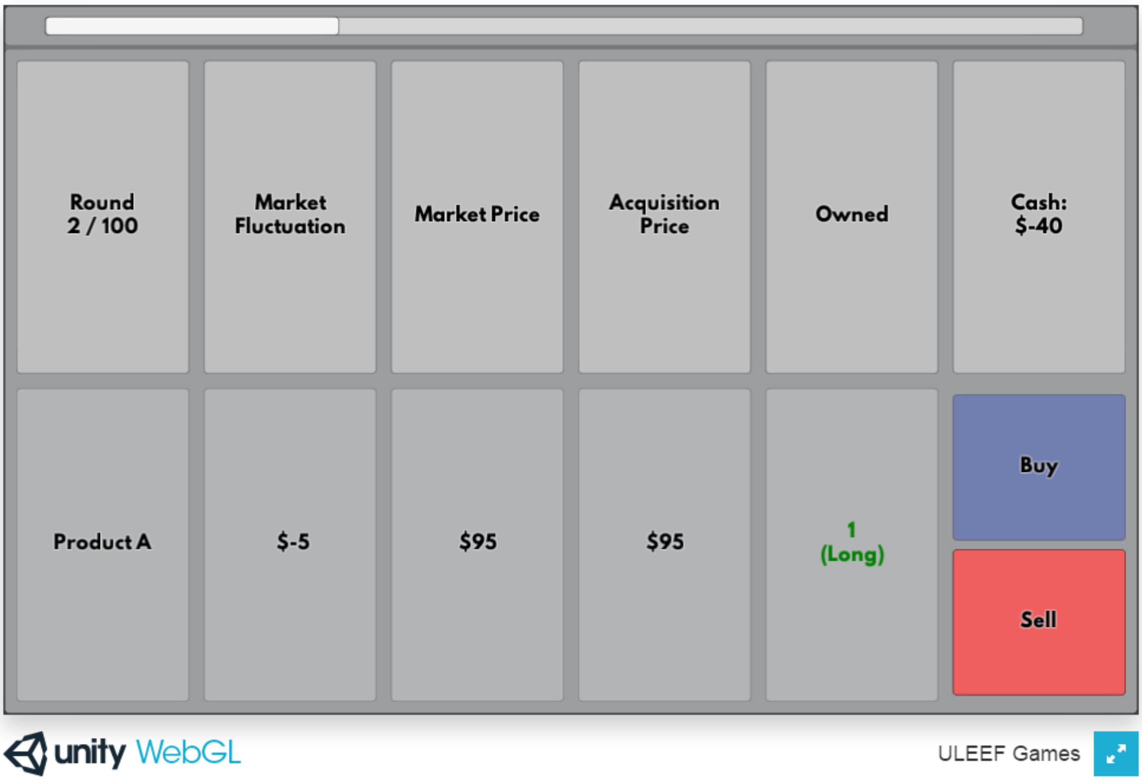
{Introductory instructions}[19]

Disposition Game

In the Disposition Game, you will attempt to maximise profits while buying and selling a security and avoiding the Disposition Effect.

Gameplay

Each round consists of the time it takes for the bar at the top of the game screen to run out (approx. 10 secs). Once the time is over, a new round is started. The timing between rounds is minimal (less than 1 second). During each round, a player can buy or sell exactly 1 security. Player's decision (submit an order to buy or sell the security, or keep the current position) is not submitted until the end of the round.



Next to the security (in this case, Product A) you will see a series of values (see above):

[19] Note, the titles in curly brackets are included for explanatory purposes only and were not seen by the participants.

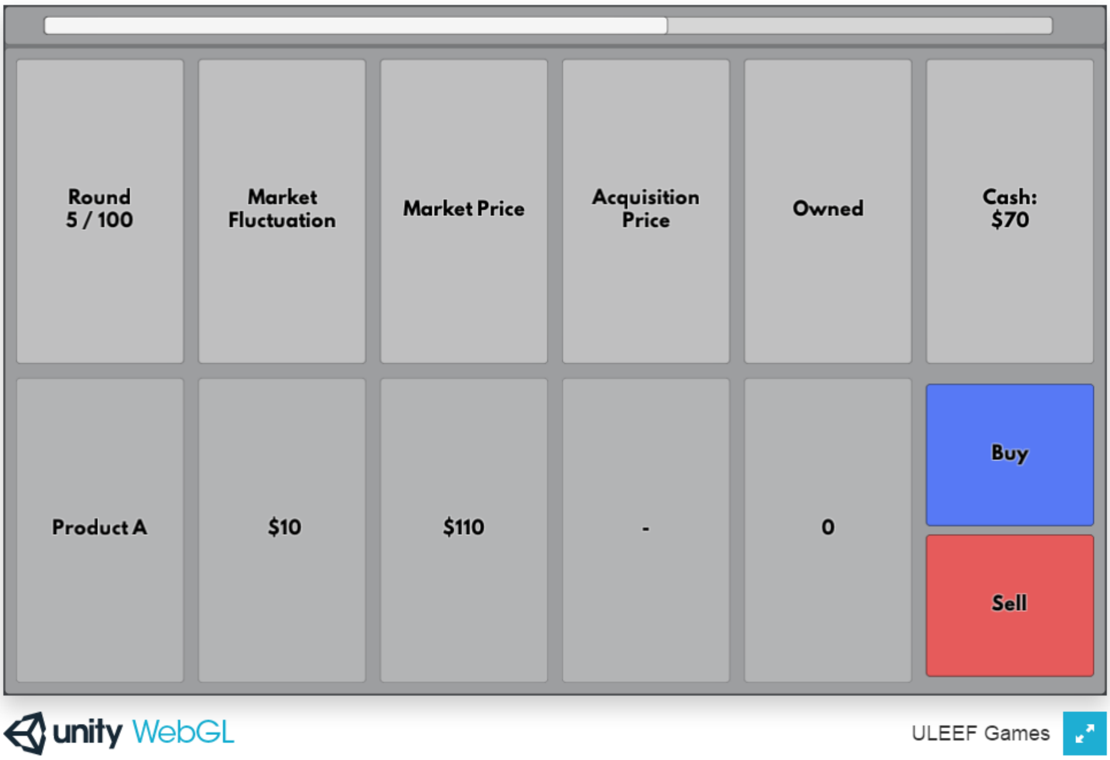
- **Market Fluctuation** is how much the market price has changed since the last round.

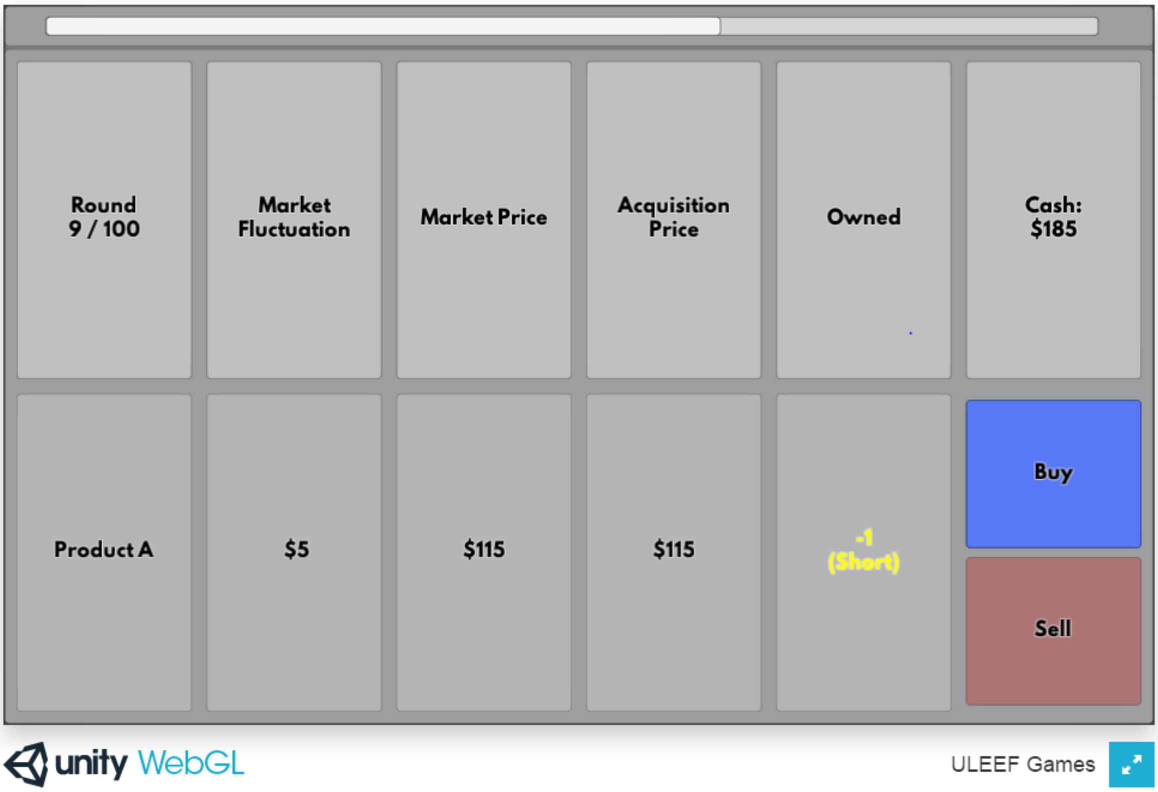
- **Market Price** is the current value of the stock if it were bought or sold.

- **Acquisition Price** is the price you purchased or short-sold the stock at last.

- **Owned** is the current amount of units of the security owned by the player.

Suppose that you are playing the Disposition Game .Once the round started and you made a decision (for example, to Buy the Product A), you may cancel your trade by pressing the “Sell” button, or you may subsequently short sell the stock by pressing “Sell” button again as it is demonstrated on two screenshots below:





The last trade before the game round finishes is considered to be your actual trade for that round. To sum up, you can undo any decision to buy or sell your security until the end of the round.

The overall game consists of four trading sessions: one practice trading session (30 rounds) to familiarise yourself with the interface of the trading game and four consecutive main trading sessions (100 rounds each), which are going to be described in more details later.

Next, you will proceed to the practice trading session.

{Instructions preceding the practice session}

You are about to begin the practice trading session. The session will consist of 30 rounds. In each round, you will have ten seconds to place an order (Buy or Sell) for one security available in the experimental market (Product A). Before the bar at the top of the game screen runs out, you can change your order. The following information is going to be provided to you on the dashboard:

- **Market Fluctuation** - how much the market price has changed since the last round.

- **Market Price** - the current value of the stock if it were bought or sold.

- **Acquisition Price** - the price you purchased or short-sold the stock at last.

- **Owned** - the current amount of units of the security owned by the player.

At the end of the practice trading session, you will be shown the amount earned for the session and the disposition effect score that characterises your investment behaviour (the higher is the value, the greater is the disposition effect). The dollar amount is provided in the end of the practice round as another indicator of how successful the selected trading strategy was. As it is just a practice trading session, this money is not going to be paid to you. However, at the end of each main trading session, the indicated dollar amount will reflect how much you actually earned throughout the session. This dollar amount will form the basis of the take-home amount that is going to be paid to you in cash at the end of today's experimental session.

For your information, at the end of the main experiment, the take-home amount will be calculated as follows. If the total value of your cash and risky asset holdings at the end of the main trading session 1 is X_1, the total value of your cash and risky asset holdings at the end of the session 2 is X_2, at the end of session 3 is X_3, and at the end of session 4 is X_4 (in experimental currency), then your take-home pay in actual dollars is equal to $10 + (X_1 + X_2 + X_3 + X_4)/48$ (in Australian dollars).

This amount will be paid to you in cash at the end of today's experiment.

Before moving forward, please raise your hand if you have any questions and the research facilitator will assist you.

{Instructions preceding the session S_1 in test treatment and session S_3 in retest treatment}

You are about to begin the first main trading session. This session repeats the design of the practice trading session, however it now consists of 100 trading rounds. Each round is 10 secs long. The amount of dollars earned (or lost) does not roll over to the following trading session(s). Thus, at the end of each trading session, your holdings of the Product A will be liquidated and the cash value of your position will be recorded. If you finish the trading session with the negative amount, you will not be required to pay this amount and your actual balance will be nullified before the next trading session.

{Instructions preceding the session O_1 in test treatment and session O_2 in retest treatment}

You are about to begin the second main trading session. In this session, you will help to make investment decisions to one of the clients of the company you are working for. Before the session begins, you will be provided with the photographs from client profiles. You will be asked to select one photograph of the client you would be helping by providing financial advice, i.e. by placing the orders for the customer. Thus, in this session you will be trading the same security, but now you will be doing it for the client. Based on your trading performance, you will be awarded a fraction (25%) of the client's earnings.

{Instructions preceding the session So_1 in test treatment and session So_2 in retest treatment}

You are about to begin the third trading session. In this session, you will have an option to place trades both for yourself and/or for your client. As in previous trading sessions, in each round you will have an option to place multiple orders before the time is up. The result of trading for yourself as well as for trading for your client (25% of the client's earnings) will be included in your reward for this trading session.

{Instructions preceding the session S_2 in test treatment and session S_4 in retest treatment}

You are about to begin the fourth and last trading session. As in the first main trading session, you are going to make investment decisions for yourself only, while trading one security. Remember, you can make multiple trading decisions within one round.

APPENDIX D: Details of clinical tests of social cognition

The Awareness of Social Inference Test - Revised (TASIT-R)[20]

The Awareness of Social Inference Test - Revised (TASIT-R) (Flanagan et al., 2011) is formed of video-recorded vignettes of typical social interactions. The test comprises three parts, each with alternate forms: the Emotion Evaluation Test (EET), the Social Inference-Minimal (SI-M) test, and the Social Inference-Enriched test (SI-E) test. EET assesses recognition of most immediate emotional expression (happy, surprised, sad, anxious, angry, disgusted, and neutral) to the social cues communicated via video vignettes. SI-M test assesses individual ability to distinguish between sincere and sarcastic cues in social interactions. Whereas the SI-E test assesses individual ability to distinguish between lies and sarcasm. In both SI-M and SI-E tests the intended meaning of the exchange is communicated via the actors' demeanour (voice and facial expression) and other contextual clues that reveal the actors' intentions. Performance on SI-E and SI-E is assessed via four standard questions per item probing for an understanding of the emotions, intentions, beliefs, and meanings of the speakers and their exchanges.

Emotion Evaluation Test (EET)

EET examines a person's ability to identify six basic emotions that are commonly recognized across cultures. The EET comprises alternative forms of a series of short (15–60 seconds) video-recorded vignettes of actors interacting in everyday situations. The professional actors, males and females, used in the EET are trained in the "Method" style, which requires the actor to elicit a real emotion in him or herself. In this way, the actors portray a spontaneous range of reactions that frequently occur with that emotion in real life. This produced a complex but more naturalistic stimulus, in comparison to the consciously derived set of cliché reactions when a person tries to "indicate" or "fake" an emotion.

[20] This overview of the tests is based on the original test overview by Flanagan et al. (2011). We recommend to check the original source for a more comprehensive overview of the tests.

The number of actors involved in each of the scenes varies. In some scenes there is only one actor (talking on the phone or directly to camera), while other scenes depict two actors and instructions are given to focus on one of them. The "target" actor in each scene enacts the script according to one of six emotions—happiness, sadness, anger, surprise, disgust (i.e., revulsion) or fear (i.e., anxiety)—or no particular emotion (i.e., "neutral"). After viewing each scene, the test participant is instructed to choose the emotion from a list of seven emotional categories. There are 28 scenes in each form, comprising four examples of each emotion in quasi-randomized order. Both forms have equal representation of scripted and improvised scenes and male and female actors. The same scripts appear on both form A and form B, but the emotion portrayed with each script differs across the two forms.

The Social Inference (SI) tests

SI tests incorporate (a) sincere verbal exchanges in which the literal meaning of the conversational remarks is sincerely meant and consistent with the physical context and emotion of the speaker; and (b) counterfactual verbal exchanges where the literal meaning of the verbal message is contradicted by the context thereby prompting viewers to derive inferences as to the true meaning of the exchange. The latter category comprises (i) lies, in which the main speaker is speaking untruthfully but wants to be believed and (ii) sarcasm, in which the main speaker makes an untrue comment that is meant to imply the opposite. In these cases, the meaning of the exchange can only be accurately interpreted by reference to such cues as the emotional state of speakers (facial expression and intonation), physical context (e.g., seeing the child's unfinished meal on the plate), and knowledge regarding who is privileged to the real facts of the situation. There are two SI tests (both with alternative forms A and B) examining the role of different types of cues in the comprehension process.

The Social Inference–Minimal (SI-M) test comprises alternative sets of 15 short (20– 60 seconds) vignettes of professional actors performing a range of scripts that represent everyday conversational exchanges that might typically occur between a couple, two friends, or two work colleagues. In five of these vignettes, the exchanges are sincere (i.e., the text and the context are consistent). For example:

(1) Michael: Sorry, I can't take that class I said I'd take on Friday.

Ruth (sincerely): That's OK, I know you're busy. Don't worry about it.

The remaining 10 vignettes encompass sarcastic exchanges where one of the speakers means the opposite of what he or she is saying and intends the recipient to understand this. In 5 of the 10 sarcastic exchanges the scripts are identical to (1) but said in such a way as to infer the opposite meaning. In the other five sarcastic vignettes the dialogue does not make sense unless it is understood that one of the participants is being sarcastic.

For example:

(2) Gary: Are you sure you've got your passport?

Keith (sarcastically): Oh, yes, I tore it up and threw it away.

Gary: Good, that's OK then.

In each of the sarcastic scenes, the cues that need to be perceived to comprehend the speaker's message include paralinguistic cues such as tone of voice and nonverbal cues such as facial expression, gesture, or body posture of the speaker (and his or her interlocutor). Failure to detect these cues will lead the viewer to interpret sincere and sarcastic versions of (1) as indistinguishable and exchanges such as (2) as meaningless or bizarre. The vignettes in both form A and form B of the SIM are edited into a quasi-random sequence.

The Social Inference–Enriched (SI-E) test comprises alternative sets of 16 short (15–60 seconds) vignettes of actors engaged in everyday conversational exchanges similar to the SI-M. In these vignettes, half represent a message that is contrary to what the main speaker believes (i.e., a lie) or contrary to the actual message the person wishes to convey (i.e., sarcasm). The lie vignettes encompass "white lies" in which the main speaker intends to deceive for the sake of diplomacy). The sarcasm vignettes involve identical scripts to lies but with cues that indicate that the speaker meant to emphasize the truth rather than conceal or minimize it.

As with part 2, the meaning of these vignettes is apparent from the paralinguistic cues provided by both the speaker and their interlocutor. In addition, part 3 vignettes provide enriched contextual cues to assist the comprehension process. Half of these are visual cues in which the viewer is provided with a camera edit showing the true state of affairs (e.g., in the example of a white lie (i) there is full). In the lie vignettes, these visual cues are apparent to the main speaker alone. In the sarcastic exchanges, both protagonists can see the visual evidence of the true state of affairs.

In the remainder of the scenes, the viewer is shown either a prologue or an epilogue to the main scene wherein one of the two main actors reveals his or her true thoughts or feelings to a third party. In all these vignettes the second main interlocutor is not privy to these asides. These cues should generally assist the comprehension process by providing additional information regarding the main speaker's knowledge and beliefs and the extent to which these are shared by the other protagonist. However, the integration of these additional sources of information demands a certain level of cognitive resources.

REFERENCES

Chiew, Kimberly Sarah, and Todd S. Braver, 2013, Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Psychology, 4* .1-15.

Flanagan, Sharon, Skye McDonald, Jennifer Rollins, 2011, *The Awareness of Social Inference Test – Revisited (TASIT-R)*. Australia: University of New South Wales, Sydney.

Frydman, Cary, Nicholas Barberis, Colin Camerer, Peter Bossaerts, and Antonio Rangel, 2014, Testing theories of investor behavior using neural data, *Journal of Finance* 69, 907–946.

Frydman, Cary, and Antonio Rangel, 2014, Debiasing the disposition effect by reducing the saliency of information about a stock's purchase price, *Journal of Economic Behavior & Organization* 107, 541–552.

Harris, Anthony D., Jessina C. McGregor, Eli N. Perencevich, Jon P. Furuno, Jingkun Zhu, Dan E. Peterson, and Joseph Finkelstein, 2006, The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical Informatics Association* 13 (1), 16-23.

Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer, 2011, *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, UK: Oxford University Press.

Komogortsev, Oleg V., Denise V. Gobert, Sampath Jayarathna, Do Hyong Koh, and Sandeep M. Gowda, 2010, Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering*, 57(11) 2635-2645.

Lacobucci, Dawn, Steven S. Posavac, Frank R. Kardes, Matthew J. Schneider, and Deidre L. Popovich, 2015, Toward a more nuanced understanding of the statistical properties of a median split, *Journal of Consumer Psychology* 25(4), 652-665.

Marchewka, A, Ł Żurawski, K Jednoróg Behavior research, 2014, The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database, *Springer*.

Motulsky, Harvey, 2013, *Intuitive biostatistics: A nonmathematical guide to statistical thinking*. New York: Oxford University Press.

Ngandu, Tiia, Jenni Lehtisalo, Alina Solomon, Esko Levälahti, Satu Ahtiluoto, Riitta Antikainen, Lars Bäckman et al., 2015, A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *The Lancet*, *385*(9984), 2255-2263.

Odean, Terrance, 1998, Are investors reluctant to realize their losses? *Journal of Finance* 53, 1775–1798.

Orquin, Jacob L., Kenneth Holmqvist, 2018, Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645-1656.

Rehder, Bob, and Aaron B. Hoffman, 2005, Thirty-something categorization results explained: selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 811.

Stampe, Dave M., 1993, Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, 25(2), 137-142.

Tabachnick, Barbara G., and Linda S. Fidell, 2007, *Experimental designs using ANOVA*. Belmont, CA: Buxbury/Thomson/Brooks/Cole.

Tobii, 2014, Tobii TX300 Eye Tracker User Manual. Tobii Technology. Available at: https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-tx300-eye-tracker-user-manual.pdf

van Rijn, Hedderik, Jelle R. Dalenberg, Jelmer P. Borst, and Simone A. Sprenger, 2012, Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. *PLoS One*, *7*(12), e51134.