**Supplementary Materials (SM)**

**Reinforcement from Sensorimotor Predictability in Major Depression**

The SM is composed of the following sections:

Participants' Medication Status

Additional analyses

Accuracy Level

Correlations of Self-Reported Measures and Mean RT

Distribution of Depression Level: HRSD & BDI

Bayesian Model Comparison

Unfiltered Results

Additional Models

Lower HRSD Cut-Point (HRSD>7)

Post hoc Testing of Psychomotor Retardation as a Predictor of RT

Discussion: our findings in light of psychomotor retardation hypotheses.

References (SI)

**Participants' Medication Status**

All participants performed the EMT task before treatment began. the total sample (N=121) included 18 participants who were on medication (~15% of the sample), their dosage was to be stable for at least three months before the start of the study and they committed to maintain that dosage for the duration of treatment. See medications list in Table S1**.**

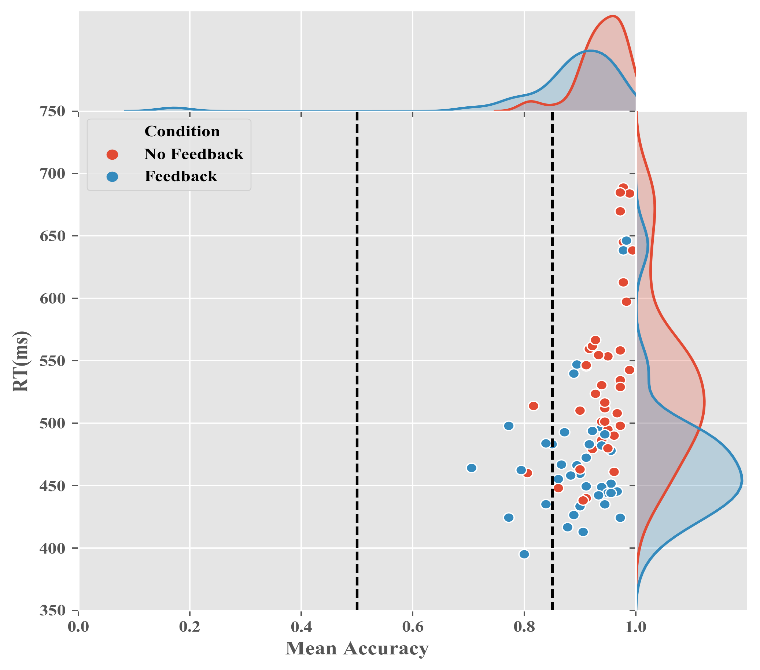
**Table S1: Summary of medications.**

|  |  |  |  |
| --- | --- | --- | --- |
| Brand-name | Generic-name | Frequency | Percent |
| - | - | 103 | 85.1 |
| Attent | Mixed Amphetamine Salts | 1 | .8 |
| Anafranil, Abilify | Clomipramine, Aripiprazole | 1 | .8 |
| Cipramil, Viepax | Citalopram | 1 | .8 |
| Viepax | Venlafaxine | 2 | 1.7 |
| Lustral, xanax | Sertralin, Alprazolam | 1 | .8 |
| Cymbalta | Duloxetine | 1 | .8 |
| Serenada | Sertraline | 3 | 2.5 |
| Parotin | Paroxetine | 1 | .8 |
| Prozac | Fluoxetine | 1 | .8 |
| Cipralex | Escitalopram | 3 | 2.5 |
| Cipramil | Citalopram | 1 | .8 |
| Concerta, Ritalin | Methylphenidate Hydrochloride | 1 | .8 |
| Remotiv | St. John Wort, Hypericum | 1 | .8 |
| Total |  | **121** | **100.0** |

**2. Additional analyses:**

**Accuracy Level**

Figure S1 depicts the screening of outlier participants regarding accuracy level.

****

**Figure S1: Scatter plot of mean accuracy of task trials (proportion), by participant. Dashed lines indicate actual filter (left) and the one used in previous experiments which used the task (right).**

**Table S1: Means (and standard deviation) for accuracy level data, by feedback.**

Subjects who did not receive RSP feedback after a key-press were, on average, significantly more accurate than those who did [t(73) = 4.41, p < .001, Cohen's d= 1.14(.65, 1.62), BF1:0 = 553.8].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | N | Mean | SD | 95% CI |
| No feedback | 38 | .95 | .03 | .94 - .96 |
| RSP-feedback | 37 | .90 | .06 | .88 - .92 |
| Total | 75 | .92 | .05 | .91 - .94 |

This finding alone may seem to point to a speed-accuracy trade-off often characterizing choice behavior. If one considers the feedback as signaling a different aspect of the task, the feedback condition may have merely influenced people to switch one goal (accuracy) for another (speed). Nevertheless, we claim this is not the case. First, as reviewed above – the factors which have been repeatedly shown to modulate the facilitation (a brief temporal lag, spatial unpredictability of the effect are consistent with a efference copy-based prediction but not with a tradeoff of goals ; e.g., Blakemore et al., 1999; Karsh et al. , 2016; Hemed et al., 2019).

Moreover, the RT facilitation effect was shown to hold even when speed is not in any explicit manner a requirement of the task (e.g., in a task to 'freely choose' a random response; Karsh & Eitam, 2015a). Finally, recent imaging work has shown that an ERP (i.e., event locked EEG activity) component that is considered to reflect the process of selection of responses is modulated in the predicted manner by the existence of own-action effects (Wen et al. , 2018).

Together we find that these findings converge to suggest that, the ‘tradeoff’ (as well as in a number of previous studies in which such a pattern was observed), is not due a simple case of (goal) tradeoff, but rather, potentially reflects the stronger and indiscriminate activation of all potential motor responses due to reinforcement (see also, Hemed et al., 2022).

**Self-Report of Perceived control**

After completing the task, participants responded to a computerized self-report questionnaire. There answers could range from 0 (not at all) to 100 (very much). The results show that there was no difference between conditions in participants' self-reports, nor an association between these measures and RT. Worth noting the lack of correlation between self-reported enjoyment (as well as perceived control and effort) and participants' performance (RT). Note that a possible explanation through the prism of hedonism and motivation in MDD cannot be rejected by the current design. First, there are documented differences between subjective reports versus objective – neural or behavioral – ‘hedonic reactions’ (see below). Objective hedonic “liking” reactions can sometimes occur alone and unconsciously without any subjective feelings of pleasure and these reactions may interact with incentive motivation in some yet unspecified manner, influence assessment of value and manipulate goal-directed behavior, despite remaining undetected by the agent itself (Berridge & Kringelbach, 2015; Childress et al., 2008; Fischman & Foltin, 1992; Morris et al., 1998; Pessiglione et al., 2007; Winkielman et al., 2005; Winkielman & Gogolushko, 2018; But see, Higgins, 1997). However, it is still unclear to what degree, if any, the RSP feedback in the current task (brief color change – from red to white – for 100 ms) can be thought of as positive-affective cue. In fact, unpublished evidence from the attempt to capture such positivity, if it exists, using an indirect measure of valence – the Affect Misattribution Procedure (AMP; Blaison et al., 2012) but no such evidence was found.

Still and this is the 2nd caveat on failing to observe a correlation between subjective hedonic feelings and RT – is the quality of measurement. Specifically, while RT is measured hundreds of times a throughout the experiment – subjective enjoyment of the task is measured once at its end. It is possible that the summary that the participant uses to respond to the enjoyment question is not sufficiently fine grained and differs from what they would have reported if enjoyment was repeatedly measured during the experiment. The problem with such a measurement is that it may effect participants’ behavior or task model and hence may (in the extreme) even produce such a correlation. Beyond that, aggregating repeated measurements *may* have reduced measurement error and with it the sizeable standard deviation of the self-reported measures. Note that this variation may reflect actual individual differences rather than measurement error.

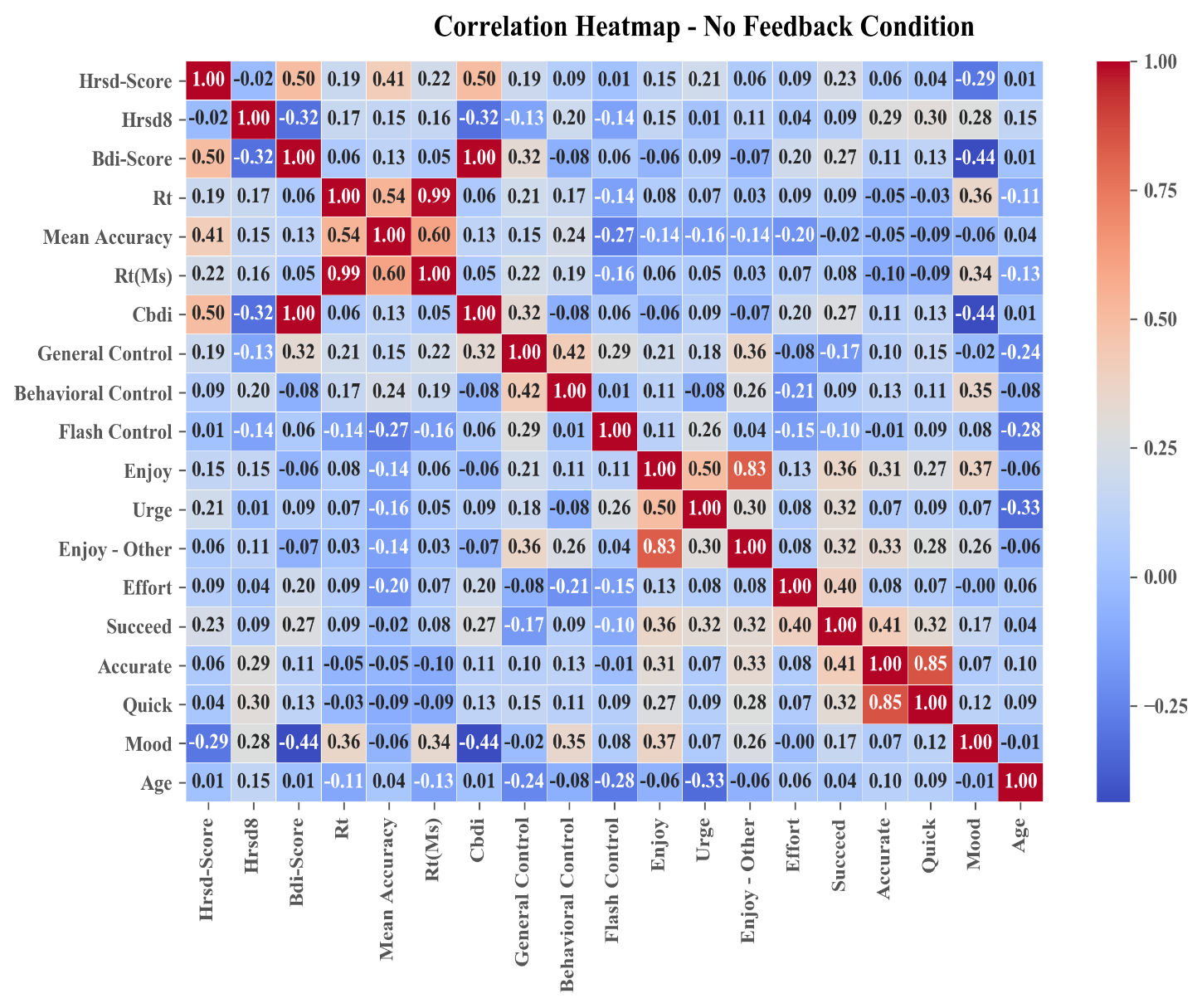
**Table S2: Summary of self-reported measures taken after completing the computerized task.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **variable** | **Range** | **Mean** | | **Standard deviation** | |
| **No-Feedback** | **RSP-Feedback** | **No-Feedback** | **RSP-Feedback** |
| **control** | General sense of control\* | 1-100 | 25.16 | 41.11 | 26.94 | 32.87 |
| sense of control on own's motor actions\* | 1-100 | 38.21 | 52.59 | 31.64 | 31.00 |
| Sense of control in the task\*\* | 1-100 | 10.26 | 84.54 | 14.09 | 20.89 |
| **Liking** | Enjoyment in the task | 1-100 | 31.42 | 39.30 | 30.79 | 28.89 |
| other's possible enjoyment in the task\*\* | 1-97 | 31.89 | 47.92 | 28.37 | 26.49 |
| **wanting** | The urge to cause a 'flash' in the task[[1]](#footnote-1)\*\* | 1-100 | 28.79 | 61.62 | 31.02 | 29.90 |
| Effort invested in the task | 1-100 | 55.34 | 57.89 | 23.69 | 26.54 |
| Intent to succeed | 1-100 | 65.82 | 72.86 | 28.61 | 25.97 |
| Intent to be accurate | 1-100 | 79.29 | 79.95 | 19.19 | 25.25 |
| Intent to be quick | 1-100 | 78.47 | 77.38 | 16.56 | 24.10 |
| **Depression level & mood** | Current mood | 1-99 | 46.68 | 42.19 | 21.83 | 21.11 |
| HRSD | 14-31 | 20.24 | 21.22 | 4.09 | 4.58 |
| BDI | 11-59 | 27.24 | 30.05 | 9.02 | 6.79 |

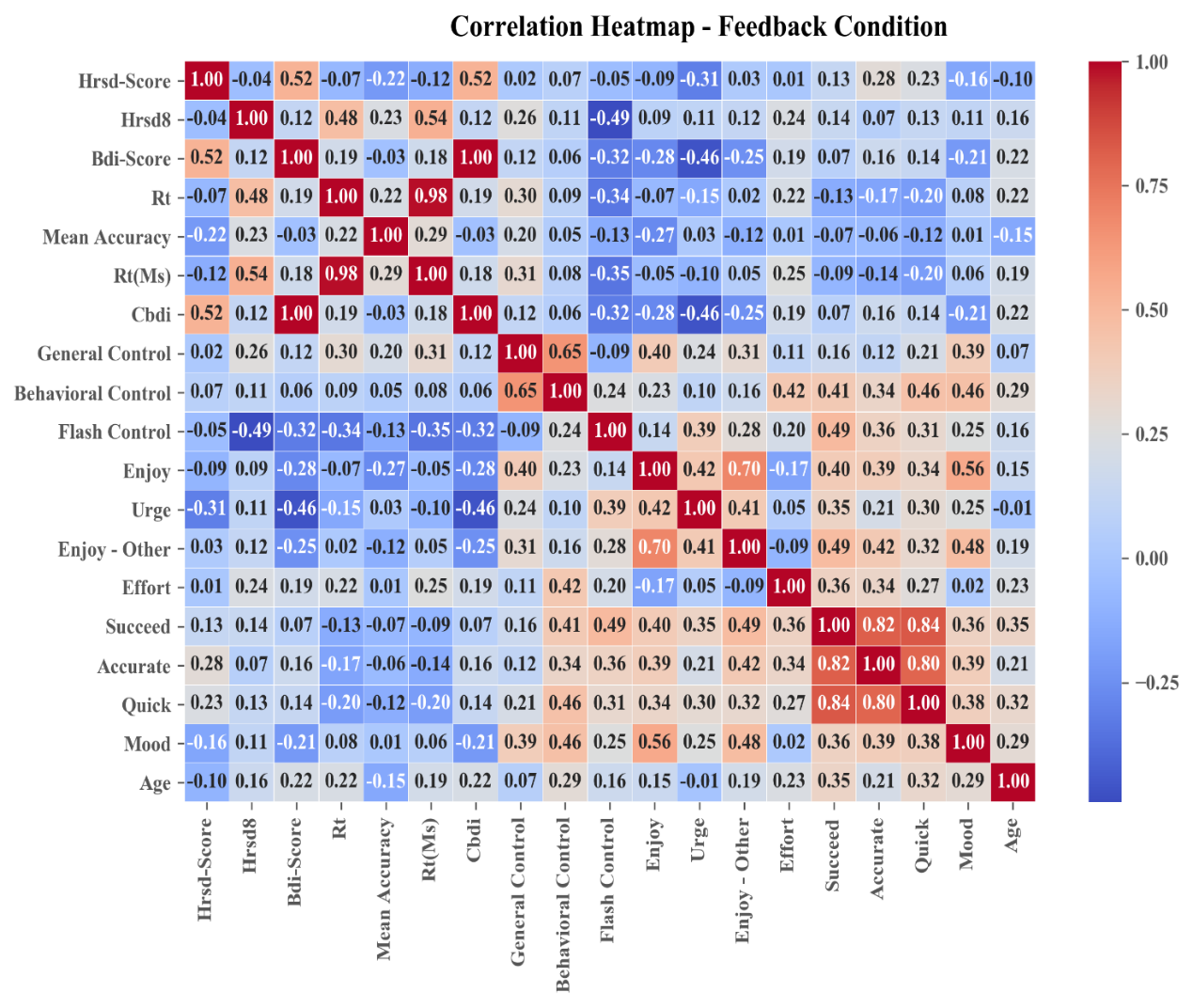
*\*significant difference between conditions, p<.05*

*\*\*significant differences between conditions, p<.01*

**Table S3: Correlation Heatmap regarding subjective measures (self-reported) and mean RT as measured in the RSP-Feedback-condition.**

****

**Table S4: Correlation Heatmap regarding subjective measures (self-reported) and mean RT as measured in the No Feedback-condition.**

****

**Distribution of Depression Level: HRSD & BDI**

HRSD scores in the No feedback condition (M=20.24, SD=4.09) were not significantly different from the scores in the RSP-feedback condition (M=21.22, SD=4.58); [t(73) = .98, p =.33, Cohen's d= -.23(-.68, .23), BF1:0 = .36].

Neither were the BDI scores in the No feedback condition (M=27.24, SD=9.02) significantly different from the scores in the RSP-feedback condition (M=30.05, SD=8.07); [t(73) = 1.52, p =.13, Cohen's d= -.35(-.81, .11), BF1:0 = .65].

**2. Bayesian Model Comparison**

In addition to the linear regression brought in the Results section of the main text we ran a corresponding Bayesian linear regression. Table S5 below contains the comparison of models produced by all the possible permutations of factors, assuming equal prior probability of each model.

The model which is best supported by the data is the one in which RSP-feedback alone (RSP in the table below) is entered into the model, with a likelihood of 0.4 and a Bayes-Factor1:0 of 2846.62, compared with the null model (containing only participant).

**Table S5: Bayesian model comparison. The most likely model is the one which includes only RSP-feedback.**

| **Model Comparison - mean\_rt** | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Models** | **P(M)** | | | **P(M|data)** | | **BF M** | | **BF 10** | | **error %** | | |
| Null model (incl. participant) |  | 0.077 |  | 1.395e -4 |  | | 0.002 |  | 1.000 |  |  |  |
| hrsd |  | 0.077 |  | 4.866e -5 |  | | 5.839e -4 |  | 0.349 |  | 1.695e -4 |  |
| bdi |  | 0.077 |  | 4.865e -5 |  | | 5.838e -4 |  | 0.349 |  | 1.695e -4 |  |
| hrsd + bdi |  | 0.077 |  | 2.035e -5 |  | | 2.442e -4 |  | 0.146 |  | 9.637e -4 |  |
| RSP |  | 0.077 |  | 0.397 |  | | 7.908 |  | 2846.615 |  | 2.214e -4 |  |
| hrsd + RSP |  | 0.077 |  | 0.126 |  | | 1.736 |  | 905.574 |  | 1.908e -4 |  |
| bdi + RSP |  | 0.077 |  | 0.151 |  | | 2.130 |  | 1080.341 |  | 2.083e -4 |  |
| hrsd + bdi + RSP |  | 0.077 |  | 0.049 |  | | 0.617 |  | 350.422 |  | 0.005 |  |
| hrsd + RSP + hrsd  ✻   RSP |  | 0.077 |  | 0.117 |  | | 1.593 |  | 839.714 |  | 0.007 |  |
| hrsd + bdi + RSP + hrsd  ✻   RSP |  | 0.077 |  | 0.045 |  | | 0.571 |  | 325.423 |  | 0.007 |  |
| bdi + RSP + bdi  ✻   RSP |  | 0.077 |  | 0.054 |  | | 0.682 |  | 385.438 |  | 0.005 |  |
| hrsd + bdi + RSP + bdi  ✻   RSP |  | 0.077 |  | 0.019 |  | | 0.237 |  | 138.664 |  | 0.002 |  |
| hrsd + bdi + RSP + hrsd  ✻   RSP + bdi  ✻   RSP |  | 0.077 |  | 0.041 |  | | 0.510 |  | 292.427 |  | 7.660e -4 |  |
|  | | | | | | | | | | | | |
| *Note.*  All models include participant. | | | | | | | | | | | | |

**3. Unfiltered Results**

In the unfiltered analysis we filtered incorrect task trials and response omissions (5.3%), but initially did not impose any other filters.Total filtration of task trials data amounted to ~9.14% and number of participants was similar (No feedback, N=40; RSP feedback, N=39).

Statistical Analysis

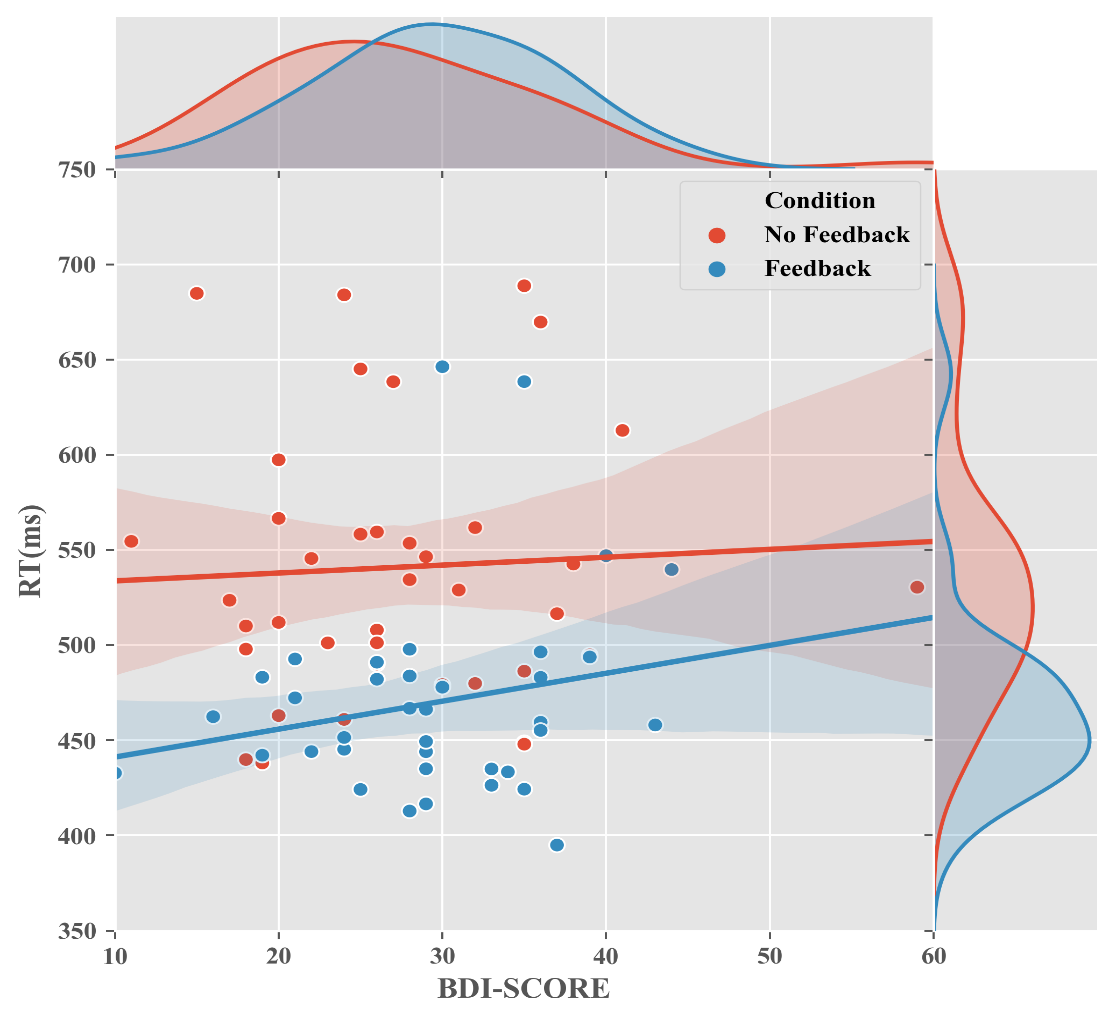
A linear regression model was used to test the influence of RSP feedback and depression levels on Reaction Time (RT). RSP feedback included two levels (feedback/no-feedback) and depression levels were represented as the participant's score on the HRSD and BDI (using two separate models as shown in Figure S2 and Figure S3).

**Figure S2: The effect of RSP-feedback and Depression level (as measured by HRSD scores) on RT.** Scatter plot indicates individual subjects' mean RT data, bands represent bootstrapped 95% CI.

The model including HRSD was significant RT (F(3,75)= 7.69, p<.001), explaining a substantial portion (20%) of the variance in. The effect of RSP-feedback on response time was significant and large (β =-.48, p>.001, BF[[2]](#footnote-2)=1203.03), which is a replication of the facilitation effect found by this paradigm in the general population (e.g., Eitam et al., 2013; Hemed et al., 2019; Karsh et al., 2016). Responses were faster in the RSP feedback condition (M=461.90, SD=74.97) compared to the No-feedback condition (M=537.17, SD=68.62).

Neither HRSD scores nor their interaction with the RSP condition reliably predicted RT [(β =.14, p=.33, BF=.23) and (β =-.14, p=.35, BF=.33) correspondingly].

The model including BDI was significant, explaining a substantial percent (20%) of the variance in RT (F(3,75)= 7.62, p<.001). The effect of RSP-feedback on response time was significant and large (β =-.48, p>.001, BF=1216.57).Neither BDI scores nor their interaction with the RSP condition reliably predicted RT [(β =.00, p=.97, BF=.24) and (β =-.09, p=.46, BF=.30) correspondingly].

**Figure S3: The effect of RSP-feedback and depression level (as measured by BDI scores) on RT.** Scatter plot indicates individual subjects' mean RT data, band represents bootstrapped 95% CI.

**4. Additional Models**

**Using a lower HRSD Cutoff for inclusion (HRSD>7 ; BDI >14)**

Results

Before analyzing the data, we used four filters based on multiple previous studies using the EMT. We rejected data in the following cases i) below accuracy-threshold; participants with more than 50%[[3]](#footnote-3) incorrect responses (~3% of the observations), ii) aberrant response times for this task: responses that were either above 800 ms or below 200 ms (0.6%), iii) incorrect responses (5~%), iv) and trials that deviated from their condition’s mean by at least 2 standard deviations in mean reaction time (~4%) or percent correct (~5%).

In addition, subjects who met the clinical criteria for depression in only one of the two clinical interviews– on the BDI-II (two participants) or the HRSD (three participants) – were also excluded from further analysis[[4]](#footnote-4).

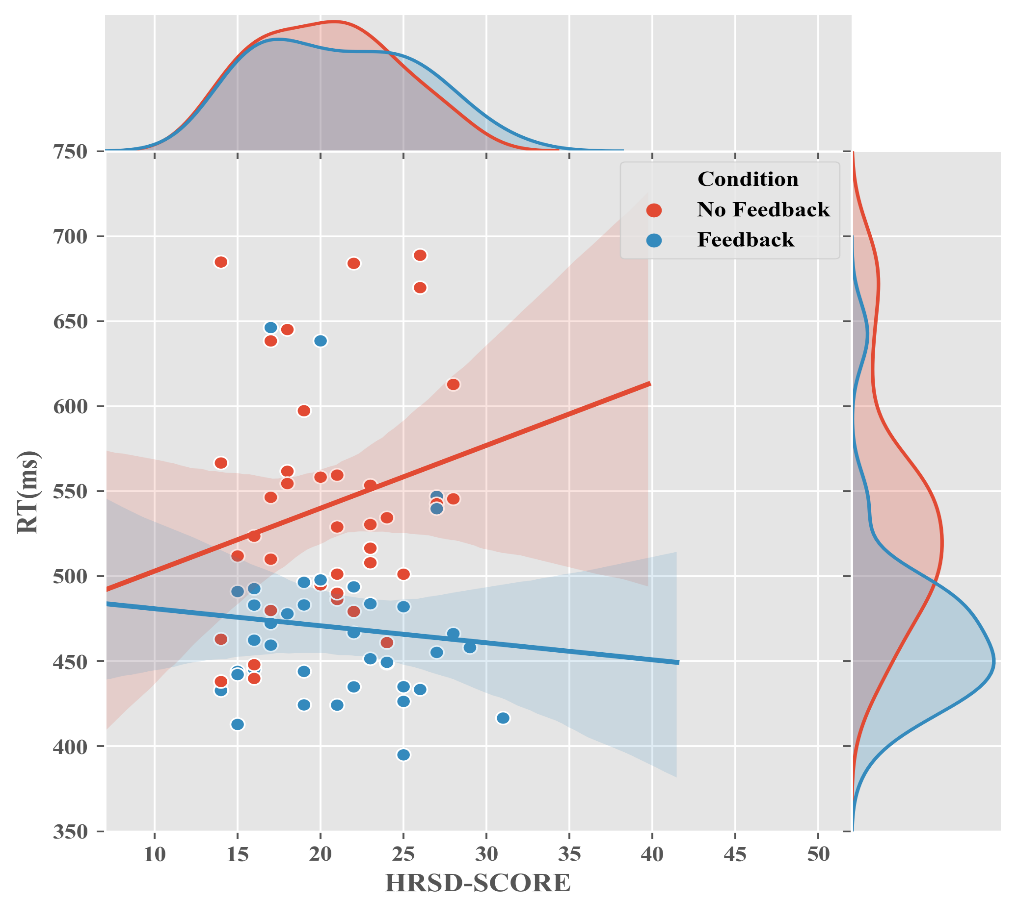
Performance: Reaction Time

A linear regression model was used to test the influence of RSP feedback and depression levels on Reaction Time (RT). RSP feedback included two levels (feedback/no-feedback) and depression levels were represented as the participant's score on the HRSD and BDI (using two separate models; Figure S3).

The model including HRSD was significant, explaining a substantial percent (21%) of the variance in RT (F(3,105)= 10.61, p<.01). Replicating previous findings obtained from the general population (Eitam et al., 2013; Hemed et al., 2019; Karsh & Eitam, 2015a; Karsh et al., 2016) we found a decisive main effect of RSP-feedback (β =-.46, p>.01, BF=9646.11). Responses were faster in the RSP feedback condition (M=540.75, SD=67.54) compared to the No-feedback condition (M=480.28, SD=55.71).

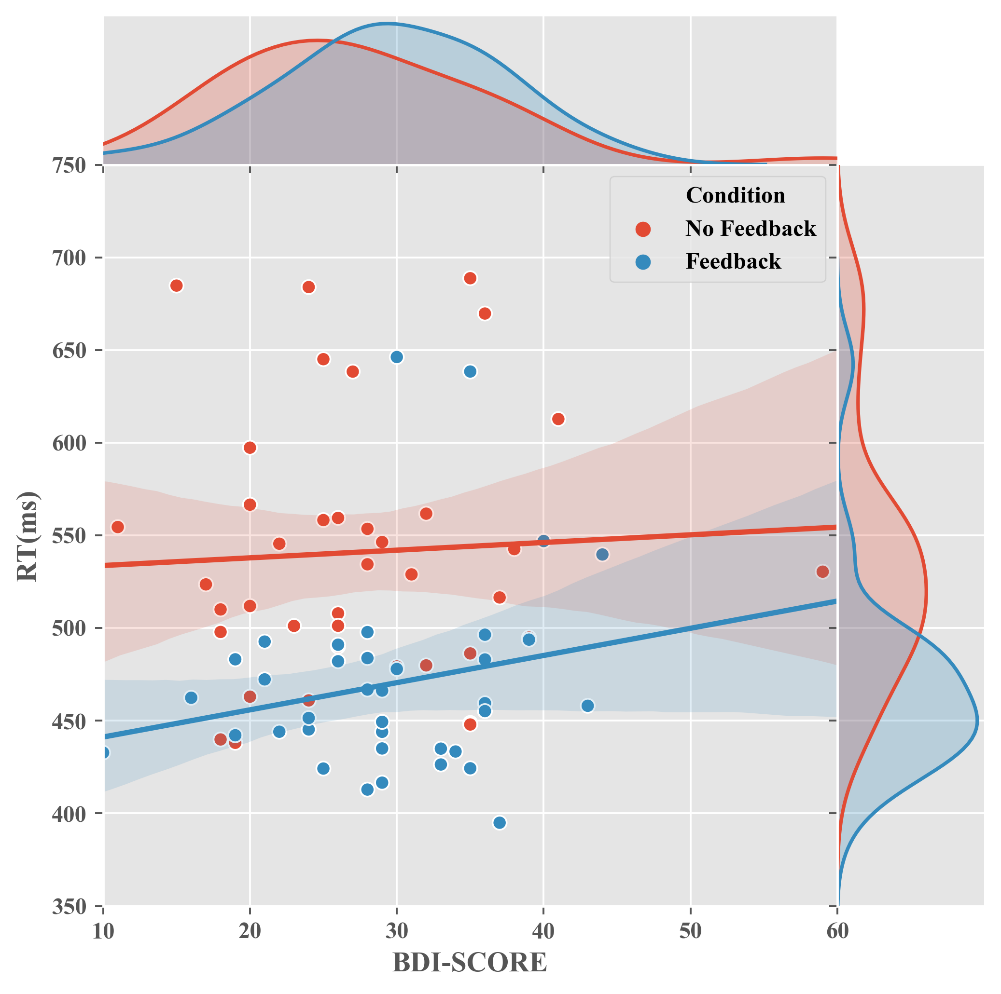
The main effect of depression level as indexed by the HRSD score was statistically significant (β =.29, p=.02, BF=.27). This seemingly indicates that a higher level of depression is associated with longer reaction times, supporting the psychomotor retardation hypothesis (Dantchev & Widlocher, 1998; Greden & Carroll, 1981; Schrijvers et al., 2008; Sobin & Sackeim, 1997); crucially – note that the Bayes factor indicates support (albeit rather weak) for the conclusion that HRSD does not affect RT.

A final pattern detected using this model, is the interaction between RSP feedback and HRSD, given its potential value (i.e. modulation of sensory motor retardation by RSP feedback) we wished to quantify the evidence for it using a Bayesian framework. although it is far from significance using a frequentist framework. The interaction’s Bayes factor was obtained directly by inserting the lower-order variables as nuisance variables in the model (which is similar to adding the interaction term following the effects of interest in a hierarchical regression), The Bayes factor for the main effects model versus the interaction term suggests that the data is insensitive to decide either for or against the existence interaction effect (β=-.22, p=.09, BF=0.90). Following additional examinations of the robustness of the interaction effect using log transformed RT and range-restriction to better capture the actual range of observed data (to avoid possible hyper-influence of few extreme observations) we downplay this potential finding that cannot be confirmed or rejected with the data we currently have.



**Figure S4: The effect of RSP-feedback and depression level (as measured by HRSD scores; cutoff HRSD>7) on RT.** Scatter plot indicates individual subjects' mean RT data, band represents bootstrapped 95% CI.

In a second model we used the person’s BDI score instead of the HRSD score for indexing their level of depression and yielded a similar pattern, explaining 18% of the variance in RT (F(3,105)= 9.09, p=<.01). With only the RSP feedback factor predicting RT (β =-.45, p=.00, BF=9646.11); neither BDI scores nor their interaction with the RSP condition were associated with RT [(β =.06, p=.61, BF=.22) and (β =.05, p=.66, BF=.27) correspondingly]. See Figure S5.



**Figure S5: The effect of RSP-feedback and depression level (as measured by BDI scores) on RT.** Scatter plot indicates individual subjects' mean RT data, band represents bootstrapped 95% CI.

**Psychomotor Retardation Item of HRSD as a Predictor of RT**

Further exploratory analyses were performed using only a single item of the HRSD that specifically measures psychomotor retardation symptoms (Item #8) as a possible predictor of RT in the EMT. One model including HRSD score after an omission of item #8 and RSP-feedback as predictors of RT. Second model including only item #8 and RSP-feedback as predictors of RT. Both models revealed no effect for the PR item (item #8) nor 'HRSD without PR item' on RT.

1st Model: HRSD without the PR item and RSP-feedback as predictors of RT yielded a similar pattern, explaining 24% of the variance in RT (F(3,71)= 8.93, p=<.01). With only the RSP feedback factor predicting RT (β =-.50, p<.001, BF=2353.71); neither 'HRSD without PR' scores nor their interaction with the RSP condition were associated with RT [(β =.20, p=.19, BF=.27) and (β =-.23, p=.13, BF=.57) correspondingly].



**Figure S6: The effect of RSP-feedback and depression level without the single item directly measuring psychomotor retardation ('HRSD without PR) on RT.** Scatter plot indicates individual subjects' mean RT data, band represents bootstrapped 95% CI.

2nd Model: PR item and RSP-feedback as predictors of RT was significant, explaining 34% of the variance in RT (F(3,71)= 13.76, p=<.01). With RSP feedback factor (β =-.47, p=.00, BF=1959.11) and PR item score (β =.22, p=.04, BF=3.92) both predicting RT. The interaction of PR item with the RSP condition was not associated with RT using a frequentist approach, yet, Bayesian approach yielded a moderate support towards the effect of PR on RT (β =.19, p=.06, BF=4.58).

However, *it should be noted that there were only 8 participants with PR item score greater than 0.* We recoded this score into a binary factor which probably made this analysis even less informative. Further work is necessary to test whether this association between PR (the observable symptoms of PR as measured in clinical interview) and RT in simple motor tasks.



RSP

**Figure S6: The effect of RSP-feedback and PR item (item #8 in the HRSD) on RT.** Scatter plot indicates individual subjects' mean RT data, band represents bootstrapped 95% CI.

**Discussion: our findings in light of psychomotor retardation hypotheses.**

Given that the mean RT in both experimental conditions and hence, the grand average of RT, was higher than that which has been documented in healthy individuals performing the same task (Eitam et al., 2013; Hemed et al., 2019; Karsh et al., 2019).

The existence of psychomotor retardation (PR) in MDD has been rather illusive to capture (Buyukdura et al., 2011), the evidence from the current study is also mixed. On one hand, the average response time of individuals suffering from MDD was higher than that documented in healthy individuals performing the same task by about 100 ms (~60%; cf. Karsh et al., 2019) and this is some evidence towards the PR. On the other hand, within the current sample of individuals with MDD, no clear association emerged between the severity of depression and RT (for neither the HRSD score or the BDI score). Hence here too, more data is required to decide either for or against the conclusion that depression level does affect RT in simple motor tasks.

Moreover, Bayes factor for the main effects model versus the interaction term suggests that the data is in fact insensitive to decide either for or against the existence of an interaction effect between depression levels and feedback condition, but only for HRSD and not for BDI. One possible explanation for this dissociation may be the different validity range of the HRSD and BDI in measuring the relevant aspects of the phenomenon. While HRSD has been reported to correlate rather highly (0.65-0.90) with other global measures of depression severity such as depressed mood, feelings of guilt and insomnia (e.g, Cole et al., 2004; Cusin et al., 2009; Miller et al., 1985; But see, Kriston & von Wolff, 2011; Santor & Coyne, 2001), the validity of the BDI has been reported to widely vary, ranging between 0.27 and 0.89 (Cusin et al., 2009).

Another explanation to the different correlation between the two measures of depression and RT may stem to the stronger emphasis on somatic and physiological complaints versus psychological symptoms in HRSD in comparison to BDI (Cole et al., 2004; Smarr, 2003), which may be more strongly associated with PR as expressed by slower RT’s.

PR is defined as a visible slowing-down of movements or speech (American Psychiatric Association, 2013). The term psychomotor itself is ambiguous because it implies both cognitive (psycho) and behavioral (motor) processes that are observed as a single act in uncontrolled settings (e.g., a clinical interview). Hence, Global ratings of PR (usually conducted by clinicians) are likely to be unreliable because they require the observer to make subjective measurements of complex behavior (Cornell et al., 1984).

However, Simple reaction time tasks are particularly suited for examining PR (Dantchev & Widlocher, 1998). Reaction time methods have been used as a more objective index of psychomotor retardation in numerous studies, and MDD groups are usually found to be slower than both healthy and nondepressed psychiatric comparison groups, on both cognitive and motor components of reaction time (e.g.; (e.g., Bezzi et al., 1981; Bruder et al., 1980; Byrne, 1976; Cornell et al., 1984).

1. Unfortunately, this difference is not very meaningful because participants in the No-feedback condition did not receive any feedback during the task – did not see any 'flash' so their urge cannot be influence by it. [↑](#footnote-ref-1)
2. We conducted Bayesian linear-regression using JASP (JASP Team, 2018) using the default Cauchy prior (width=0.707); Conventionally, Bayes factors between 3 and 10 are considered moderate evidence for the test hypothesis (null or alternative), and a Bayes factor greater than 10 is considered strong evidence. When .3<BF<3, BF evidence is considered inconclusive (Jeffreys, 1961; Wagenmakers et al., 2018). [↑](#footnote-ref-2)
3. In previous experiments conducted on healthy individuals, accuracy threshold was set to 85% (Hemed et al., 2019; Karsh & Eitam, 2015a, 2015b;Karsh et al., 2016), however due to our concern that very slow reactions would also be coded as no-response (and by so – as an incorrect one) and our wish to maximize the number of observations, we decided to lower the threshold. Key results of the model do not change when no filters are applied. [↑](#footnote-ref-3)
4. After applying the filters, the number of participants in both conditions was similar (N=55 in No feedback condition and N=54 in RSP Feedback condition) [↑](#footnote-ref-4)