Elevated effort cost identified by computational modeling as a distinctive feature explaining multiple behaviors in patients with depression

Fabien Vinckier, MD-PhD, Claire Jaffre, Claire Gauthier, MD, Sarah Smajda, MD, Pierre Abdel-Ahad, MD, Raphaël Le Bouc, MD-PhD, Jean Daunizeau, PhD, Mylène Fefeu, MD, Nicolas Borderies, PhD, Marion Plaze, MD-PhD, Raphaël Gaillard, MD-PhD, Mathias Pessiglione, PhD

PII: S2451-9022(22)00184-7
DOI: https://doi.org/10.1016/j.bpsc.2022.07.011
Reference: BPSC 980

To appear in: Biological Psychiatry: Cognitive Neuroscience and Neuroimaging

Received Date: 19 August 2021
Revised Date: 14 July 2022
Accepted Date: 25 July 2022

Please cite this article as: Vinckier F., Jaffre C., Gauthier C., Smajda S., Abdel-Ahad P., Le Bouc R., Daunizeau J., Fefeu M., Borderies N., Plaze M., Gaillard R. & Pessiglione M., Elevated effort cost identified by computational modeling as a distinctive feature explaining multiple behaviors in patients with depression, Biological Psychiatry: Cognitive Neuroscience and Neuroimaging (2022), doi: https://doi.org/10.1016/j.bpsc.2022.07.011.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Inc on behalf of Society of Biological Psychiatry.
Elevated effort cost identified by computational modeling as a distinctive feature explaining multiple behaviors in patients with depression

Short title (55 characters): Effort cost in depression

Authors:

Fabien Vinckier¹,²,³* (MD-PhD), Claire Jaffre¹,²,³, Claire Gauthier²,³ (MD), Sarah Smajda²,³ (MD), Pierre Abdel-Ahad²,³ (MD), Raphaël Le Bouc¹,⁴,⁵ (MD-PhD), Jean Daunizeau¹,⁶ (PhD), Mylène Fefeu²,³ (MD), Nicolas Borderies² (PhD), Marion Plaze²,³ (MD-PhD), Raphaël Gaillard²,³,⁷, # (MD-PhD), and Mathias Pessiglione¹,⁶, # (PhD)

Affiliations:

¹ Motivation, Brain & Behavior (MBB) lab, Institut du Cerveau (ICM), Hôpital Pitié-Salpêtrière, F-75013, Paris, France

² Université Paris Cité, F-75006 Paris, France

³ Department of Psychiatry, Service Hospitalo-Universitaire, GHU Paris Psychiatrie & Neurosciences, F-75014 Paris, France

⁴ Urgences cérébro-vasculaires, Pitié-Salpêtrière hospital, Sorbonne University, Assistance Publique - Hôpitaux de Paris (APHP), Paris, France.

⁵ Zurich Center for Neuroeconomics, Department of Economics, University of Zurich, Zurich, Switzerland.

⁶ Sorbonne Universités, Inserm, CNRS, Paris, France

⁷ Institut Pasteur, experimental neuropathology unit, Paris, France

* Corresponding author

# these authors contributed equally to the work

To whom correspondence should be addressed: Fabien Vinckier, Service Hospitalo-Universitaire, GHU Paris Psychiatrie & Neurosciences, 1 rue Cabanis, 75014 Paris
Tel: (0033) 1 45 65 84 52 ; Fax: (0033) 1 45 65 81 60
mailto:fabien.vinckier@u-paris.fr
Keywords: depression, reward, punishment, effort, decision making, computational modeling

Abstract

Background: Motivational deficit is a core clinical manifestation of depression and a strong predictor of treatment failure. However, the underlying mechanisms, which cannot be accessed through questionnaire-based conventional scoring, remain largely unknown. According to decision theory, apathy could result either from biased subjective estimates (of action costs or outcomes) or from dysfunctional processes (in making decisions or allocating resources).

Methods: Here, we combined a series of behavioral tasks with computational modeling to elucidate the motivational deficits of 35 patients with unipolar or bipolar depression under various treatments, compared to 35 matched healthy controls.

Results: The most striking feature, observed independently of medication across preference tasks (likeability ratings and binary decisions), performance tasks (physical and mental effort exertion) and instrumental learning tasks (updating choices to maximize outcomes), was an elevated sensitivity to effort cost. By contrast, sensitivity to action outcomes (reward and punishment) and task-specific processes were relatively spared.

Conclusions: These results highlight effort cost as a critical dimension that might explain multiple behavioral changes in patients with depression. More generally, they validate a test battery for computational phenotyping of motivational states, which could orient toward specific medication or rehabilitation therapy, and thereby help pave the way to a more personalized medicine in psychiatry.
Brief summary: Motivational deficit is a core clinical manifestation of depression. We combined a series of behavioral tasks with computational modeling to elucidate the motivational deficits of 35 patients with depression, compared to 35 healthy controls. The most striking feature, observed across different tasks assessing both performance and preference, was an elevated sensitivity to effort cost. These results highlight effort cost as a critical dimension that might explain multiple behavioral changes in depression.
Introduction

“Nature has placed mankind under the governance of two sovereign masters, pain and pleasure”(1). In the line of this famous statement from Jeremy Bentham, mood is classically conceived as “oscillating between the two extremes of pleasure and pain”(2). Consistently, standard descriptions of mood disorders such as depression focused on the “psychic pain”(3) experienced by patients, or their “anhedonia”(4), i.e. their inability to experience pleasure. However, for Bentham, a pioneer of utilitarian decision theory, pain and pleasure are masters in the sense that they drive the behavior: people essentially strive to enjoy pleasure and to avoid suffering pain. In this regard, patients with depression are not normally driven - their motivational deficit has recently emerged as a pathological cornerstone of depression. A key reason is that motivational deficit is one of the best predictors of functional impairment and subjective quality of life in depression(5, 6). Another is that motivational deficit remains less responsive to conventional treatment than standard mood-related symptoms. For instance, the interest-activity dimension of clinical questionnaires is a strong predictor of poor response to antidepressants, above and beyond depression severity(7). Also, motivational deficit is frequently reported as a residual symptom after adequate treatment by serotonergic antidepressants in unipolar depression, or mood stabilizers in bipolar disorder(8). Finally, motivational deficit could contribute to other dimensions of depression such as executive dysfunction and account for reduced efficiency in cognitive tests(9).

Yet the way motivational deficit is assessed in current textbooks or clinical questionnaires does not align with modern decision theory. In this theory, the agent is supposed to engage in the action that maximizes a cost/benefit trade-off. The benefits relate to the outcome of the action, i.e. obtaining a reward or avoiding a punishment. The costs may relate to the action itself, such as effort, or modulate the value of the outcome, such as delay.
In this view, items such as ‘reduced activity’ or ‘concentration difficulty’ (lack of engagement) would be the consequence of either ‘low energy’ (higher expected effort cost) or ‘low interest’ (lower expected reward value). Indeed, reduction of goal-directed behavior could result either from a decreased sensitivity to outcomes (I do nothing because I see no purpose in potential activities) or from an increased sensitivity to effort (I do nothing because the costs of actions are too high, even if the pleasures and pains at stake are still meaningful to me)(10, 11).

In this view, effort is thus an attribute of actions that should be distinguished from punishment (or loss), which is like reward (or gain) an attribute of outcomes. This distinction was overseen by Bentham himself, who listed effort among the nine pains of the senses, insisting on the “uneasy sensation which is apt to accompany any intense effort, whether of mind or body.” Even today, effort stricto sensu is virtually absent from modern definitions of depression, although it is sometimes alluded to through the notions of fatigue or lack of energy. However, as stated in the DSM-5, patients with depression often report that “even the smallest tasks seem to require substantial effort”, while “the efficiency with which tasks are accomplished may be reduced”(12). These descriptions suggest that effort cost is increased in depression, leading patients to either have a more aversive sensation if they invest the same effort as healthy people, or to be less efficient than healthy people if they match the aversive sensation by investing less effort.

The aim of the present study is to properly dissociate the impact of depression on the sensitivity to effort, punishment and reward, as conceptualized in decision theory. Note that we opted for the words reward / punishment because they designate outcomes of actions, rather than gain / loss which refer to changes in wealth than can be passively experienced (like with lotteries). To go beyond what clinical questionnaires can tell(13), and assess the integrity of motivational control processes, we set up a battery of behavioral tasks. This is important,
not only because questionnaires do not exactly assess the dimensions that are key to behavioral control, but also because they heavily rely on the quality of insight. In addition, behavioral tests present the advantage that they can be paralleled in animal models, opening an avenue for more invasive investigations of the neurophysiological mechanisms that might dysfunction in patients.

In brief, two main kinds of behavioral tasks have been used to assess motivational impairment in depression. One line of research focused on reward versus punishment processing, typically using reinforcement learning paradigms(14). Classical results in depression suggest a reduced sensitivity to reward(15-17), which has been linked to anhedonia and dopaminergic transmission(18). Results regarding sensitivity to punishments are less consistent, with studies showing worse performance following negative outcomes and others blunted responses to negative stimuli(19-24). Another line of research focused on the effort/reward trade-off. The effort dimension was first introduced with an incentive motivation test assessing the force exerted on a handgrip device as a function of the money at stake(25). Since this seminal paper, reduced willingness to exert effort for reward has been reported in many studies, using different behavioral read-outs (binary choice, willingness to engage effort, effort-dependent performance) and different kinds of efforts (key pressing, handgrip squeezing, cognitive control tasks)(26-29) in various clinical populations (subsyndromal vs. actual depression, unipolar vs. bipolar depression, and drug-naïve patients(28, 29)). A reward/effort trade-off task has also been recently used to predict relapse after antidepressant discontinuation(30).

While they made an important breakthrough, these studies suffer from limitations. First, they typically employ a unique behavioral task, taking the risk that results may depend on some specific task features that may not generalize across different contexts (e.g. the
nature of reward, usually money, or the mode of response, usually choice). Second, they typically contrast positive and negative dimensions, such as reward versus punishment or reward versus effort, failing to assess whether two negative dimensions such as effort and punishment are differentially affected in depression. Third, they seldom use mechanistic models that would help pinpoint the covert dysfunctional process, e.g., whether a reduced willingness to work is due to decreased sensitivity to reward or increased sensitivity to effort.

To this aim, a promising approach consists in phenotyping motivation states by fitting computational models to patients’ behavior(31, 32). Crucially, this computational approach can be used to discriminate between several cognitive dysfunctions that may result in a similar overt behavioral deficit(33), hence bridging the gap between clinical assessment and the underlying pathophysiology(34, 35).

Using this approach, we assessed the behavior of patients with depression (n=35, MDE) and matched healthy controls (n=35, HC) in a comprehensive battery of preference, performance and learning tasks that involve two types of outcome (reward and punishment) and two types of cost (effort and delay). The behavior of patients and controls were then compared using computational models that tracked dysfunctional processing of a same motivational factor across different tasks.
Methods & Materials

Participants

The study was approved by local Ethics Committee (CPP Ile de France 3, Paris, France). A total of 70 participants, including 35 patients with depression and 35 healthy controls completed the study. All participants were informed that they would not be paid for their voluntary participation and that the monetary earnings in the task were purely fictive.

Patients were recruited in inpatient and outpatient facilities. They all met criteria for major depressive episode, with a MADRS score >20 and a background diagnosis of either bipolar disorder or major depressive disorder (single or recurrent episode). Patients with a diagnosis of schizoaffective disorder were excluded. Healthy participants were recruited from the community. All participants were native French speakers, had normal or corrected-to-normal vision and gave informed consent before taking part. Patients and controls had no history of brain injury, epilepsy, alcohol or other drug abuse, or neurological disorders. Healthy controls were also screened for any history of psychiatric conditions, psychoactive substance abuse or dependence, or psychotropic medication. There was no significant difference between patients with depression (MDE group) and healthy controls (HC group) regarding age, gender, or education level (see table 1).

Table 1: Demographic and psychometric details

<table>
<thead>
<tr>
<th></th>
<th>MDE (N = 35)</th>
<th>HC (N = 35)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (F / M)</td>
<td>18 / 17</td>
<td>18 / 17</td>
<td>1</td>
</tr>
<tr>
<td>Age (years)</td>
<td>42.5 (16.9)</td>
<td>43 (16.3)</td>
<td>.902</td>
</tr>
<tr>
<td>Education (years)</td>
<td>15.2 (3.8)</td>
<td>15.1 (2.5)</td>
<td>.881</td>
</tr>
</tbody>
</table>
Currently active* (N) & 21 & 31 & .116 \\
PCSA (cm²) & 43.0 (10.6) & 43.2 (10.0) & .929 \\
MADRS (depression) & 36.1 (7.4) & 2.9 (2.8) & < 0.001 \\
Starkstein (apathy) & 23.8 (5.6) & 7.4 (4.0) & < 0.001 \\
SHAPS (anhedonia) & 5.4 (4.3) & 0.5 (1.0) & < 0.001 \\
Thase & Rush staging & 6 / 15 / 9 / 5 & NA \\

Data provided in cells are numbers or means (sd). MADRS: Montgomery - Asberg depression rating scale. SHAPS: Snaith - Hamilton pleasure scale. PCSA: physiological cross-sectional area (morphological proxy for maximal force, see supplementary materials).

* Including patients in sick leave and students

The MDE group included both patients with bipolar disorder (N = 15) and patients with major depressive disorder (N = 20, single or recurrent episode). Note that the two subgroups are not matched and that our sample is underpowered to assess any difference between diagnoses. We nevertheless report demographic, psychometric and treatment details separately for the two conditions (see table 2).

### Table 2: Comparison of patient with MDD and bipolar disorder

<table>
<thead>
<tr>
<th></th>
<th>MDD (N = 20)</th>
<th>Bipolar disorder (N = 15)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (F / M)</td>
<td>11 / 9</td>
<td>7 / 8</td>
<td>0.625</td>
</tr>
<tr>
<td>Age (years)</td>
<td>45.7 (19.0)</td>
<td>38.1 (12.9)</td>
<td>0.192</td>
</tr>
<tr>
<td>Education (years)</td>
<td>14.8 (4.1)</td>
<td>15.8 (3.2)</td>
<td>0.444</td>
</tr>
<tr>
<td>Currently active* (N)</td>
<td>11</td>
<td>9</td>
<td>0.767</td>
</tr>
<tr>
<td>MADRS (depression)</td>
<td>37.2 (6.9)</td>
<td>34.7 (8.1)</td>
<td>0.333</td>
</tr>
<tr>
<td>Starkstein (apathy)</td>
<td>23.7 (6.3)</td>
<td>24.0 (4.7)</td>
<td>0.858</td>
</tr>
<tr>
<td>SHAPS (anhedonia)</td>
<td>5.6 (4.7)</td>
<td>5.1 (3.8)</td>
<td>0.723</td>
</tr>
<tr>
<td>Thase &amp; Rush staging</td>
<td>2 / 7 / 6 / 5</td>
<td>4 / 8 / 3 / 0</td>
<td>0.104</td>
</tr>
<tr>
<td>Lithium (N)</td>
<td>0</td>
<td>6</td>
<td>0.002</td>
</tr>
<tr>
<td>Medicine Type</td>
<td>N</td>
<td>Mean</td>
<td>p-value</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>Antidepressant (N)</td>
<td>20</td>
<td>4</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Atypical antipsychotic (N)</td>
<td>5</td>
<td>10</td>
<td>0.013</td>
</tr>
<tr>
<td>Anxiolytic (N)</td>
<td>9</td>
<td>8</td>
<td>0.625</td>
</tr>
</tbody>
</table>

Data provided in cells are numbers or means (sd). MADRS: Montgomery - Asberg depression rating scale. SHAPS: Snaith - Hamilton pleasure scale.

* Including patients in sick leave and students

Experiment

All details about behavioral tasks and computational models can be found in supplementary methods.
Results

All participants performed preference, performance and learning blocks of tasks in this order. For each block (preference, performance or learning tasks), we only report model-based analyses in the main text, the model-free analyses (as well as additional model-based analyses) can be found in supplementary results. Note that with our sample size, only moderate to large effects (Cohen’s D > .68) could be detected at standard statistical thresholds (power of 80% and significance level at 5%).

Preference tasks

This block contained 4 tasks that presented the same natural items belonging to one of three dimensions: reward, punishment and effort (see Fig.1). Reward items could be food (e.g. ‘to get a chocolate cookie’) or goods (e.g. ‘to get a standard 32-card deck’), punishment items could be sensory (e.g. ‘to hear a chalkboard screech’) or more abstract (e.g. ‘to have my phone screen scratched’), while effort items could be physical (e.g. ‘to walk up 5 floors of stairs’) or mental (e.g. ‘to fill in an administrative form’). Items were presented as short texts, except for an extra-set of reward items that were accompanied by illustrative pictures.

In the likeability rating task (Fig.1A), participants were instructed to rate how much they would like to be given the reward (likeability rating of appetitive items) or how much they would dislike being imposed the punishment or effort (dislikeability rating of aversive items). In the binary choice task (Fig.1B), participants were asked to select their favorite item among two options belonging to the same dimension (i.e. the reward they prefer to obtain or the punishment / effort they prefer to endure). In the yes/no choice subtask (Fig.1C), they were asked to state whether they would accept or decline a hypothetical option representing
a trade-off between two dimensions (exerting an effort to obtain a reward, exerting an effort to avoid a punishment, enduring a punishment to obtain a reward). Finally, in the inter-temporal choice subtask (Fig.1D), they were asked to state their preference between two hypothetical options combining two dimensions, i.e. an item associated with a delay. Thus, they had to choose between a small reward (or punishment or effort) implemented immediately and a more likeable (or less dislikeable) one later in time.

Model-based analyses

The standard analysis of choice tasks consists in using a softmax function that transforms option values into selection likelihood, which is equivalent to logistic regression (Fig.1). When options are natural items, with no explicit attribute like magnitude and probability of monetary gain or loss, option values are given by likeability ratings. In the case of a binary choice between A and B, the probability Proba_A of selecting option A (against B) is given by the softmax function:

$$\text{Proba}_A = \text{sig}(\beta \cdot (\text{Rating}_A - \text{Rating}_B)),$$

where \text{Proba}_A is the probability of selecting option A, \text{Rating}_A is the likeability rating of item A, and \text{Rating}_B is the likeability rating of item B. The function \text{sig}(x) = 1 / (1 + \exp(-x)) is the logistic function, and \beta is a free parameter termed “inverse temperature”, which captures the consistency of choices: higher \beta means less stochastic choices. Typically, the purpose of fitting such choice model is to test whether ratings are good predictors of choices. However, this was not our case: we aimed at inferring the subjective values that patients assigned to the different items, given both likeability ratings and choices. In that regard, the standard analysis is heavily biased: likeability ratings are assumed to be exact (noiseless) expression of subjective values, whereas choices are assumed to be stochastic (noisy) expression of those subjective values. A fair approach would consider both ratings and choices as noisy expressions of the same
underlying hidden values\(^{(36)}\). Thus, our general approach was to fit all tasks together, with a unique model, to extract these hidden values. Specifically, the subjective value of each item was represented by one free parameter. These hidden-value parameters were then mapped through task-specific observation functions to produce the behavioral response (rating or choice). These functions contained other free parameters, on top of hidden values, which were specific to the task but common to all dimensions. Thus, dimension-specific differences between groups could not be captured by task-specific parameters but only by the distribution of hidden values across the relevant set of items.

When fitted to all behavioral data concurrently, this computational approach provides estimates of two types of free parameters: hidden-values parameters (one per item) and task-specific parameters (weights, biases, and discount factors) that control the mapping from hidden values to behavioral responses (choices and ratings). For each participant and dimension, model inversion provided a distribution of hidden-value parameters. The summary statistics (mean and variance) of value distributions (over items) were taken as measures of each participant’s sensitivity to the different dimensions (reward, effort, punishment).

*Model-based Results*

Observed data and model fits for all tasks are illustrated in Fig.2. The distribution of hidden values across items and participants are illustrated in Fig.3A. The subject-wise mean hidden-value was entered in ANOVA with dimension as a within-subject factor, group (patients with major depressive episode or healthy controls, respectively labeled MDE and HC groups in the following) as a between-subject factor and subject as a random factor (Fig.3B). There was a main effect of dimension (\(p<0.001\)), indicating that our items were not balanced (punishments were more aversive than efforts). Beyond this, there was a main effect of group (\(F(1, 68) =\)
and an interaction between group and dimension (F(2, 136) = 3.8; p = 0.025).
Post-hoc t-tests revealed that this interaction was mainly driven by higher (i.e. more aversive)
hidden values for effort items (μE) in the MDE group compared to HC group (5.3 vs. 4.4; t(68) = 3.2; p = 0.002), while there was no difference between groups for all other dimensions.
Higher hidden values in the MDE group were observed both for motor (5.9 vs. 4.8) and
cognitive efforts (4.8 vs. 3.9), both p<0.05. The same ANOVA was conducted on subject-wise
standard deviation (across the set of items) and task-specific free parameters, see
supplementary results and supplementary Fig.1 for details).
Thus, the main computational feature that distinguished MDE patients from controls
was an increased aversion for effort cost (μE), which captured higher ratings of effort items
and higher tendency to avoid effort in yes/no choices (see description of model-free analyses
in supplementary results).

Performance tasks
The above preference tasks suggest that patients with MDE exhibit a higher sensitivity to
effort when compared to healthy controls. However, these tasks are purely declarative (since
all options are hypothetical). To assess whether the same elevated effort sensitivity would
manifest when the effort is not virtual and has to be exerted, we tested participants on
“performance” tasks that were previously used in fMRI and clinical studies [37, 38]. In both
motor and cognitive performance tasks (Fig.4), participants exerted effort so as to maximize
their (virtual) monetary payoff. Both tasks included 10 series of 12 trials in which participants
played either to maximize monetary earnings or to minimize losses (on previously earned
money). Each trial was associated with one of 6 possible monetary incentives (€0.01, €0.2,
€0.5, €1, €5, and €20), corresponding to coins and notes used in everyday life in France. In the
motor performance task, participants squeezed a handgrip, while in the cognitive performance task, they performed a series of numerical comparisons between digits displayed with different font size, generating a Stroop effect. Participants were instructed that the payoff was proportional to both the incentive at stake and their performance (peak of force pulse in the grip task or number of correct responses in the Stroop task).

**Model-based analyses**

We used the same model to fit raw performance measures in the grip and Stroop tasks, namely peak force (in Newtons) and correct response rate (number per second), respectively. This model was already applied to motor performance in a previous study that specified the computational phenotype of motivational deficit in patients with Parkinson’s disease (37) and was here extended to cognitive performance (see Fig. 4, supplementary methods and supplementary Fig. 2 for details).

**Model-based Results**

Observed data and model fits for all tasks are illustrated in Fig. 5.

When comparing parameter estimates between groups (Fig. 6), the most striking difference was a higher $K_c$ (sensitivity to effort cost) in patients relative to controls, for both motor performance ($0.108$ vs. $0.028$, $t(68) = 6.0$, $p < 0.001$) and cognitive performance ($0.014$ vs. $0.009$, $t(68) = 2.8$, $p = 0.006$). There was also a trend for lower $K_i$ (sensitivity to incentive value) but it was not significant in motor performance ($0.743$ vs. $1.001$, $t(68) = -1.4$, $p = 0.172$) and bordering significant in cognitive performance ($0.004$ vs. $0.012$, $t(68) = -2.0$, $p = 0.054$). Finally, $P_{\text{max}}$ was lower in patients than in controls, for both the grip ($326$ vs. $443$N, $t(68) = -$
2.8, \( p = 0.006 \) and the Stroop task (1.71 vs. 2.3 correct responses / second, \( Z = -4.1, p < 0.001 \)). There was no significant difference between groups in \( K_f \) (fatigue effect).

Thus, the main computational feature that distinguished MDE patients from controls was an increased aversion for effort cost (\( K_e \)), which captured a globally lower motor and cognitive performance (see description of model-free analyses in supplementary results).

Learning task

In the last block, participants performed three sessions of the same instrumental learning task, previously used in fMRI and pharmacological studies to dissociate between positive and negative reinforcement (by reward and punishment). We observed no such dissociation, as and the main distinction between groups was about choice stochasticity (see supplementary material for detailed results).

Link with clinical factors

Clinical scores – depression severity (MADRS), apathy (Starkstein’s scale), and anhedonia (SHAPS) – were regressed against a GLM containing mean hidden values for reward and effort items (\( \mu_R \) and \( \mu_E \)) in preference tasks. There was a significant link between the effort parameter and the apathy score, in the sense that more aversive effort value predicts more severe apathy (\( \beta = -0.36, p = 0.032 \)). However, this association requires confirmation in larger samples, since it would not survive correction for multiple comparisons if considering all possible links between computational parameters and clinical scores (there was no other significant link). Similarly, correlations between tasks (e.g. between parameters fitted to preference vs. performance tasks), failed to reach significance. In particular, there was no
significant correlation between mean value for effort items and the weight of (motor or cognitive) effort cost (both \( p > 0.1 \)).

Our sample of patients was diverse in terms of diagnosis (unipolar vs. bipolar disorder) and medication (antidepressants / anxiolytics / atypical antipsychotics / lithium) and too small for conducting direct comparisons between subgroups. However, the main results (elevated mean effort value \( \mu_E \) and effort cost \( K_E \) relative to controls) were observed in all subgroups, whatever the diagnosis and medication (see Fig. 7 and supplementary results for details). Thus, they were not driven by a particular subgroup and can be generalized across different types and treatments of major depressive episode.
Discussion

In this paper, we used a series of behavioral tasks, coupled with computational modeling, to assess motivational deficits in patients with depression compared to matched healthy controls. More specifically, we assessed three key dimensions of motivational control that, in principle, could result in a reduction of goal-directed behavior: increased sensitivity to action costs (effort) and decreased sensitivity to action outcomes (reward and punishment). Even if we observed some reduced sensitivity to action outcomes, the most striking result was a massive increase in the sensitivity to effort cost, observed both in the preference tasks (mean aversive value of effort items, $\mu_e$) and in the performance tasks (weight of effort cost on net expected value, $K_e$). In the following, we discuss the interpretation of computational parameters that were altered in our patients, the possible links between our results and other conceptual frames for motivational control, and the clinical implication of these results regarding pathological mechanisms and therapeutic interventions.

The typical approach in cases versus controls studies of cognitive impairment is to use one behavioral task to assess one cognitive process. However, different tasks might be sensitive to different facets of the target process and present peculiarities that might bias the assessment (39, 40). To better ensure that results would be generalizable across contexts, we followed a conjunction method, looking for similarities rather than differences between tasks. In addition, we employed a computational approach that distinguished task-specific parameters from the variables of interest that represent similar cognitive construct across tasks. Indeed, patients expressed an increased aversion for effort costs that was manifest in both preference and performance tasks. This is important because the two sets of tasks have different strengths and weaknesses. Performance tasks present the advantage of assessing the effort that patients really invest for goal-directed behavior (not just a declared intention
to exert effort). They manipulate potential rewards in a systematic manner and provide objective performance measures, which indirectly quantify the effort invested. They have been extensively used to show a reduced willingness to exert effort for reward in patients with depression (25-29). Here, we replicate and extend this typical result by isolating the contribution of effort cost estimates from the impact of expected rewards. A difficulty is to discriminate between possibilities of more aversive effort and more limited capacity (i.e., ‘not willing to’ versus ‘not being able to’). In other terms, patients’ reduced performance could be due to motor or cognitive capacity loss, rather and not because doing the task properly would feel too effortful. Our strategy was to dissociate parameters that estimate maximal possible performance (i.e., capacity) from parameters that scales the expected cost to the expected benefit in the net value function that determines effort allocation (see supplementary material for a discussion devoted to this specific issue).

An increase in the aversiveness of both motor and cognitive effort cost was also observed in preference tasks. Compared to performance tasks, these rating and choice tasks present the advantage of directly assessing subjective effort costs. They also offer the possibility to generalize potential deficits to many different reward, punishment or effort items that are faced in real life, and not just motor and cognitive tasks made in the lab to win or avoid losing money. Thus, the observed enhanced effort sensitivity across preference and performance tasks provides strong evidence that this unique alteration explains both what patients feel (and report) and what they actually do when faced with a motor or cognitive challenge. However, we did not find a between-task correlation (across patients) in our measures of effort sensitivity ($\mu_e$ and $K_e$). Although this may be due to limited sample size, it could also suggest that subjective effort costs in declared intentions (preference tasks) and in effective trade-offs (performance tasks) are partially dissociable (41).
Contrary to effort sensitivity, reward and punishment sensitivity did not differ much between patients and controls. Notably, there was no difference in the mean values of reward and punishment items ($\mu_R$ and $\mu_P$) in preference tasks. The enhanced effort sensitivity, while outcome sensitivity is relatively preserved, is reminiscent of the preserved liking versus impaired wanting that has been observed in many conditions, including depression and anhedonia\cite{27, 42, 43}. Our tasks did not enable the systematic comparison of motivation and consumption phases that is required to specifically test the liking / wanting dissociation. However, our findings may shed a new light on this dissociation, which was originally tested in rodents\cite{44} by comparing the affective reactions to reward delivery (liking) and the willingness to exert effort for obtaining reward (wanting). Thus, at the heart of “wanting” is the trade-off between reward and effort that we have tested in our performance tasks. Under this view, our results provide complementary evidence for the notion that, at least in some cases, impaired wanting could be simply reduced to excessive sensitivity to effort costs.

Even if decreased outcome sensitivity (in patients versus controls) was sometimes bordering significance, it was always similar for reward and punishment, whether the task was about preference, performance or learning. This is an important result because many studies have emphasized that low mood might lead to overweighting negative stimuli, relative to positive stimuli\cite{45-47}. Typical results involve excessive processing of negative stimuli and higher sensitivity to negative feedback, possibly leading to learned helplessness, and a lower sensitivity to positive feedback, associated with anhedonia\cite{14, 22, 48}. Our results rather single out effort cost as the distorted negative dimension (that shifts downward the expected net values driving actions) in depression. Comparatively, patients’ processing of the other negative dimension (punishment) was not altered. In other words, the critical line of divide between depressed and non-depressed people may not be between positive and negative.
events, but between action costs and outcomes (irrespective of whether they are anticipated or experienced).

Obviously, we do not claim that effort sensitivity is the only dimension affected in all patients suffering from depression. This claim cannot be derived from our data, first because it would be based on null results, and in any case because our results are limited to the factors tested in the behavioral tasks and to a small sample of patients that may not be representative of depression in general. However, our results do suggest that sensitivity to effort cost is significantly more reliably affected than sensitivity to the other dimensions tested, which is important because previous studies using reward / effort tradeoff tasks and models did not distinguish between these two possible explanations of apathy in depression - rewards feeling less desirable and effort feeling more exhausting.

A limitation of our study was the absence of strong correlation between the key computational markers of depression and the clinical dimensions assessed with questionnaires. This is in line with previous studies that failed to show correlation between severity of depression and willingness to produce effort for reward(28, 29). This suggests that computational modeling of behavior brings complementary information to standard questionnaires. Of course, this may not be the case, should clinical scales used to score depression include more questions about how aversive effort is for patients in their daily life. Also, we may not have the statistical power to test the correlation between clinical scores and computational parameters, as our sample was modest and mainly composed of severely depressed patients (most of them being hospitalized when tested). The diversity in our sample may be seen as a weakness for drawing strong conclusions about a specific clinical feature, but also as a strength for a better generalization of the results, which is not possible when inclusion criteria are so narrow that all patients fall in the same subgroup. Indeed, we
leveraged this diversity to show that our main findings (elevated aversion for effort relative to controls) were significant even when restricting the analysis to patients with a particular diagnosis or medication. Further studies may be needed to confirm our results in a more homogenous sample of unmedicated patients.

Finally, computational phenotyping of motivational deficits may help bridge the gap between clinical assessment of depression and its underlying neurobiology. The link between dopaminergic transmission and sensitivity to reward has been largely studied (49, 50), including in reward / effort trade-off tasks (27, 37, 51-53), while punishment and sensitivity to losses have been linked with opioid transmission (54). The pharmacology of effort has been seldom explored (compared to pain and reward), but a recent study demonstrated that classical antidepressant (citalopram) might help overcome effort cost (55), which would be consistent with the use of serotoninergic medication in the treatment of depression. But if motivational deficits in depressed patients are due to a salient increase in effort cost, then why are they weakly improved by serotoninergic antidepressant treatment? A simple answer could be that patients’ effort sensitivity is too high to be normalized by serotoninergic medication. Alternatively, a minority of patients may suffer from other deficits on top of increased effort cost, including alterations of reward processing, which was shown to condition the response to antidepressant treatment (56, 57). In the latter case, effort sensitivity would be a marker of the disease, while reward sensitivity would be a marker of drug resistance. Given the link between reward processing and dopaminergic transmission, then individual patients with reduced reward sensitivity (on top of the common enhanced effort sensitivity) might be better treated with drugs that combine serotoninergic with dopaminergic actions. Under this perspective, computational phenotyping of patients’ motivational state would help predict which treatment should be used in each patient.
Acknowledgment

The study was funded by Fondation Bettencourt Schueller. CG was funded by Fondation pour la Recherche Médicale (FRM). The authors are grateful to Lionel Rigoux, Pablo Carrillo, and Lucie Berkovitch for helpful insight.

Disclosure

FV has been invited to scientific meetings, consulted and/or served as speaker and received compensation by Lundbeck, Servier, Recordati, Janssen, Otsuka, LivaNova, and Chiesi. He has received research support by Lundbeck. CG has served as speaker and received compensation by Pileje. SS has been invited to scientific meetings by Janssen, Otsuka, Lundbeck, and Servier. PAA has been invited to scientific meeting by Actelion and Otsuka. CJ consulted and received compensation by Janssen and Lundbeck. MP has served as speaker and received compensation by Lundbeck, Janssen, and LivaNova. RG has received compensation as a member of the scientific advisory board of Janssen, Lundbeck, Roche, SOBI, Takeda. He has served as consultant and/or speaker for Astra Zeneca, Boehringer-Ingelheim, Pierre Fabre, Lilly, Lundbeck, LVMH, MAPREG, Novartis, Otsuka, Pileje, SANOFI, Servier and received compensation, and he has received research support from Servier. None of these links of interest are related to this work. The remaining authors report no biomedical financial interests or potential conflicts of interest.
Legends

**Fig.1 Preference block: task and modeling principles.**

Reward, punishment, and effort natural items were used in four tasks. A. Rating task: participants were required to rate how appetitive rewards were (likeability rating) or how aversive punishments and efforts were (dislikeability rating). B. Binary choice task: participants were required to choose between two items the one they prefer. C. Yes/no choice task: participants were required to hypothetically accept or decline an option combining two dimensions. D. Inter-temporal choice task: participants were required to make hypothetical choices between two items, one to be experienced immediately and one (more likeable or less dislikeable) after a delay. All behavioral outputs (ratings and choices) were fitted together, using free parameters representing the hidden values ($V_{item}$, in pink) of the different items across tasks, and free parameters (slope and bias, in blue) adjusting the sigmoid mapping specific to each task.

**Fig.2 Preference block: model-free results and model fits.**

Observed data (lines) and model fits (diamonds) are shown for each of the four preference tasks. A. Rating task: observed and modeled rating as a function of hidden value (inferred through model fitting) for reward (left panel), punishment (middle panel), and effort (right panel). B. Binary choice task: observed and modeled choice rate (probability of choosing left) as a function of the difference between left and right item value for reward (left panel), punishment (middle panel), and effort (right panel). C. Yes/no choice task: observed and modeled choice rate (acceptance probability) as a function of the difference between the hidden values of benefit (obtained reward or avoided punishment) and cost (exerted effort or
inflicted punishment) for reward/effort (left panel), punishment/effort (middle panel), and reward/punishment (right panel) trade-off. D. Intertemporal choice task: observed and modeled choice rate (probability of choosing the immediate option) as a function of the difference between hidden values of immediate and delayed items (using exponential discounting) for reward (left panel), punishment (middle panel), and effort (right panel). Shadows represent inter-subject s.e.m.

**Fig.3 Preference block: model-based results.**

Stacked histograms of hidden values for both groups (depressed patients and healthy controls on top and bottom panels) and all dimensions (reward, punishment and effort on left, middle and right panels). Each participant is represented with a different (arbitrary) color. B. Mean hidden values for each dimension. Difference between groups was significant for the effort dimension only. White line: median; box 25th (Q1) and 75th (Q3) percentiles of the distribution over the group; Points: outlier participants for whom the mean is larger than Q3+1.5 * (Q3-Q1) or smaller than Q1-1.5 * (Q3-Q1); Whiskers min to max (without outliers).

**Fig.4 Performance block: task and modeling principles.**

In these performance tasks, participants had to produce an effort (either motor or cognitive) so as to maximize their (virtual) monetary payoff. Each trial was associated with one of 6 possible monetary incentives (€0.01, €0.2, €0.5, €1, €5, and €20). In both tasks, a visual feedback on current performance level was provided as a cursor moving up a scale. A. In the motor performance task, participants had to squeeze a handgrip as hard as possible within a 3-s time window. The top of the scale corresponded to maximal force produced during calibration. B. In the cognitive performance task, participant had to perform as many as possible numerical comparisons between digits displayed with different font size, generating
a Stroop effect. The time allowed was set to 70% of the time taken during calibration to complete all 10 numerical comparisons. Payoff was proportional to both the incentive at stake and their performance (peak of force pulse in the grip task or number of correct responses in the Stroop task). Participants played either to win some money (as illustrated for the motor task in A) or to not lose previously earned money (as illustrated for the cognitive task in B). The feedback screen indicated both the money won or lost in the current trial and a cumulative total. C. Cost-benefit optimization model. Simulated hidden variables. Perf (left part) is a saturation function linking resource spent (u) to performance F(u). F(u) tends to $P_{\text{max}}$ (theoretical maximal performance) when u tends to $\infty$, without inflexion point. The expected net value (right part) of possible resource investment u at a given trial t is obtained by subtracting costs from benefits (see main text and supplementary Fig.2 for details). The optimal resource $u^*$ and resulting perf and net value are illustrated in the graph for four out of the six incentive levels. Free parameters are written in blue.

**Fig.5 Performance block: model-free results and model fits.**

Observed raw performance (unnormalized) and modeled performance (diamonds) as a function of incentive level in the motor performance task (peak of force pulse in Newton, left panel) and in the cognitive performance task (rate of correct responses, right panel). Lines are means and shadows represent inter-subject s.e.m.

**Fig.6 Performance block: model-based results.**

Summary statistics of the free parameters fitted on motor (A) and cognitive (B) performance. Difference between groups was significant for $K_c$ and $P_{\text{max}}$ (in both tasks) and marginal for $K_i$ in the cognitive performance task. White line: median; box 25th (Q1) and 75th (Q3) percentiles of the distribution over the group; Points: outlier participants for whom the standard-
deviation is larger than Q3+1.5 * (Q3-Q1) or smaller than Q1-1.5 * (Q3-Q1); Whiskers: min to max (without outliers).

Fig. 7 Effect of the diversity of patients and treatments on effort parameters

Summary statistics of $\mu_E$ and $K_c$ (in both tasks) as a function of diagnoses and treatments (boxplot of patients minus the average of HC). For each parameter, we represent the whole MDE group (medium blue) and each subgroup separately (light and dark blue). For most factors of diversity, the difference with controls followed the same trend in the two subgroups (with and without the considered factor) and was significant in most cases. None of the direct comparisons between subgroups was significant. White line: median; box 25th (Q1) and 75th (Q3) percentiles of the distribution over the group; Points: outlier participants for whom the standard-deviation is larger than Q3+1.5 * (Q3-Q1) or smaller than Q1-1.5 * (Q3-Q1); Whiskers: min to max (without outliers).
Reference


A. How aversive would it be to climb 5 floors

Not at all

Extremely

How pleasant would it be to get a cookie

Not at all

Extremely

rating_cookie = \text{sig} (\beta_r \times V_{\text{cookie}} + \beta_{mr})

ingrating_climb = \text{sig} (\beta_r \times V_{\text{climb}} + \beta_{mr})

B. What do you prefer?

To get an apple

To get a cookie

C. I accept to climb 5 floors to get a cookie

Yes

No

D. What do you prefer?

To get an apple now

To get a cookie in 1 day

\text{proba_left} = \text{sig} (\beta_{c1} \times (V_{\text{apple}} - V_{\text{cookie}}) + \beta_{mc})

\text{proba_left} = \text{sig} (\beta_{c2} \times (V_{\text{cookie}} - V_{\text{climb}}) + \beta_{mc})

\text{proba_left} = \text{sig} (\beta_{c3} \times (V_{\text{apple}} - V_{\text{cookie}}) \times e^{-kd} \times \text{Delay} + \beta_{mc})
\[ \text{NetValue}(u, t) = \frac{\text{Perf}(u)}{P_{\text{max}}} + \frac{\text{Perf}(u)}{P_{\text{cal}}} \times \text{Inc}(t) - K_c \times (1 + K_f \times \frac{1}{N_t}) \times u^2 \]

\[ \text{Perf}^* = \text{Perf}(u^*) = \text{Perf}(u) \left( \arg\max(\text{NetValue}) \right) \]