Research Paper

Greater decision uncertainty characterizes a transdiagnostic patient sample during approach-avoidance conflict: a computational modelling approach

Ryan Smith, PhD; Namik Kirlic, PhD; Jennifer L. Stewart, PhD; James Touthang, BS; Rayus Kuplicki, PhD; Sahib S. Khalsa, MD, PhD; Justin Feinstein, PhD; Martin P. Paulus, MD; Robin L. Aupperle, PhD

Background: Imbalances in approach-avoidance conflict (AAC) decision-making (e.g., sacrificing rewards to avoid negative outcomes) are considered central to multiple psychiatric disorders. We used computational modelling to examine 2 factors that are often not distinguished in descriptive analyses of AAC: decision uncertainty and sensitivity to negative outcomes versus rewards (emotional conflict). **Methods:** A previously validated AAC task was completed by 478 participants, including healthy controls (n = 59), people with substance use disorders (n = 159) and people with depression and/or anxiety disorders who did not have substance use disorders (n = 260). Using an active inference model, we estimated individual-level values for a model parameter that reflected decision uncertainty and another that reflected emotional conflict. We also repeated analyses in a subsample (59 healthy controls, 161 people with depression and/or anxiety disorders, 56 people with substance use disorders) that was propensity-matched for age and general intelligence. **Results:** The model showed high accuracy (72%). As further validation, parameters correlated with reaction times and self-reported task motivations in expected directions. The emotional conflict parameter further correlated with self-reported anxiety during the task (r = 0.32, p < 0.001), and the decision uncertainty parameter correlated with self-reported difficulty making decisions (r = 0.45, p < 0.001). Compared to healthy controls, people with depression and/or anxiety disorders and people with substance use disorders showed higher decision uncertainty in the propensity-matched sample (t = 2.16, p = 0.03, and t = 2.88, p = 0.005, respectively), with analogous results in the full sample; people with substance use disorders also showed lower emotional conflict in the full sample (t = 3.17, p = 0.002). **Limitations:** This study was limited by heterogeneity of the clinical sample and an inability to examine learning. **Conclusion:** These results suggest that reduced confidence in how to act, rather than increased emotional conflict, may explain maladaptive approach-avoidance behaviours in people with psychiatric disorders.

Introduction

Imbalances in the decision to approach or avoid when both positive and negative consequences are expected (i.e., approach-avoidance conflict; AAC) is often problematic in people with mental health conditions.[1] For example, people with depression and anxiety may choose to sacrifice participation in rewarding activities because they believe such activities will also lead to negative consequences (e.g., judgment, embarrassment[2]). People with substance use disorders also engage in costly drug-taking behaviours to avoid negative affect and show impairment during decision-making in tasks that involve conflict between reward and punishment (reviewed in Ekhtiari and colleagues[3] and Guttman and colleagues[4]). A better understanding of the underlying factors that contribute to avoidance-based decision-making may improve our understanding of these mental health conditions and inform the development of treatments that target the distinct factors that are relevant for individual patients.[5]

Several paradigms are used to study AAC (for a review, see Kirlic and colleagues[6]), most of which create conflict between receiving monetary rewards and either monetary punishments (monetary-based conflict[7,8]), pain (pain-based conflict[9]) or aversive affective stimuli (affect-based conflict[10–13]).

Correspondence to: R. Smith, Laureate Institute for Brain Research, 6655 S. Yale Ave., Tulsa, OK 74136, USA; rsmith@laureateinstitute.org

Submitted Feb. 13, 2020; Revised May 15, 2020; Accepted Jun. 22, 2020; Early-released Oct. 29, 2020

DOI: 10.1503/jpn.200032

Although AAC is often analyzed using more traditional behavioural measures, computational modelling has emerged as a promising new approach for analysis.[14–19] Modelling allows for precise quantification of distinct information-processing mechanisms that contribute to decision-making. For example, in a monetary-based conflict study,[20] variations in a model parameter that reflected relative sensitivity to reward versus punishment accounted for sex differences in avoidance (greater avoidance behaviour in females), whereas greater sensitivity to punishment overall better accounted for the increased avoidance behaviour observed in those with an inhibited temperament. In a pain-based conflict study,[9] computational modelling also showed that sensitivity to reward-based benefits became increasingly attenuated as pain-based costs increased in aversiveness.

Computational modelling has not yet been applied to affect-based conflict. Doing so may add clinical value, because anticipation of negative affect may be particularly important for understanding the factors that contribute to psychiatric disorders. To be clear, the distinction made here between affect-based conflict and pain- or monetary-based conflict is not meant to suggest that losing money or feeling the unpleasant sensation of pain do not involve affective responses.[21] The distinction motivating the use of affect-based tasks is one of maximizing ecological validity: affective disorders more often involve avoidance of the more complex "emotional pain" induced by socio-emotional cues (e.g., social rejection, losing a job, death). The visual/auditory stimuli used in affect-based AAC tasks are designed to more closely match such cues (e.g., by depicting social interactions mirroring some of those cues) and the emotionally painful responses they evoke.

In this study, we applied a computational modelling approach[22,23] to study affect-based conflict. We demonstrated how this approach could separate 2 underlying components of conflict that have not been thoroughly distinguished in traditional descriptive analyses: decision uncertainty and relative sensitivity to negative affective stimuli versus reward (emotional conflict). Both uncertainty and emotional conflict have potential relevance for psychiatric disorders. For example, poor decision-making in people with anxiety has been associated with high intolerance of uncertainty and risk aversion,[24,25] whereas suboptimal decision-making in depression appears to be driven more by attenuated responses to reward.[26,27] Substance use disorders have also been associated with intolerance of uncertainty[28] and a reduced ability to incorporate uncertainty into reward learning during decision-making tasks.[29] However, the distinct contributions of uncertainty versus emotional conflict (threat/reward sensitivity) to avoidance behaviour have not been fully delineated.

Using a relatively large sample comprising healthy controls without psychiatric symptomatology and a transdiagnostic sample of patients with depression, anxiety and/or a substance use disorder, we estimated model parameters that reflected decision uncertainty and emotional conflict using an affect-based AAC task.[11,12,30] We hypothesized that relative to healthy controls, both psychiatric groups would exhibit greater uncertainty and greater emotional conflict in the AAC task.

Methods

Participants

We identified participants for this analysis from the first 500 participants in the Tulsa 1000 (T1000),[31] a naturalistic longitudinal study that recruited participants based on the dimensional National Institute of Mental Health Research Domain Criteria framework.[32] The T1000 study included a community-based sample of approximately 1000 people recruited through radio, electronic media, treatment centre referrals and word of mouth (this sample size was planned a priori; see Victor and colleagues[31] for a detailed justification based on the aims of the larger study). Participants were 18 to 55 years of age and screened on the basis of dimensional psychopathology scores: Patient Health Questionnaire (PHQ-9[33]) score ≥ 10; Overall Anxiety Severity and Impairment Scale (OASIS[34]) score ≥ 8; and/or Drug Abuse Screening Test (DAST-10[35]) score > 2. The healthy controls showed no elevated symptoms or psychiatric diagnoses. Participants were excluded if they tested positive for drugs of abuse; met the criteria for a psychotic disorder, bipolar disorder or obsessive–compulsive disorder; reported a history of a moderate to severe traumatic brain injury, a neurologic disorder or a severe or unstable medical condition; reported an active suicidal intent or plan; or reported a change in medication dose within 6 weeks of participation in the study. Full inclusion and exclusion criteria are described in Victor and colleagues.[31] The study was approved by the Western Institutional Review Board. All participants provided written informed consent before they completed the study protocol, in accordance with the Declaration of Helsinki, and they were compensated for participation (ClinicalTrials.gov identifier NCT02450240). A number of previous papers have been published from the larger T1000 data set,[36–46] but none of these papers has included analyses or data from the AAC task.

Given the heterogeneous clinical sample in the T1000 and its explicitly transdiagnostic focus, we divided participants into 3 groups: healthy controls; people with substance use disorders; and people with depression and/or anxiety who did not have substance use disorders. Participants were grouped based on DSM-IV or DSM-5 diagnosis using the Mini International Neuropsychiatric Inventory,[47] and analyses focused on groups of participants with major depressive and/or anxiety disorders (social anxiety, generalized anxiety, panic and/or posttraumatic stress disorder; n = 260); substance use disorders (recreational drugs excluding alcohol and nicotine, with or without comorbid depression and/or anxiety disorders; n = 159); and healthy controls with no mental health diagnoses (n = 59).

As further described in Victor and colleagues,[31] the T1000 study was built around the National Institute of Mental Health Research Domain Criteria framework, which describes dimensions of pathology.[32] Thus, the T1000 study specifically focused in advance on recruiting participants with these symptom profiles, with the aim of identifying transdiagnostic behavioural and neural phenotypes that were related to threat/reward processing, interoceptive processing

and cognitive functioning. Although symptoms can be observed dimensionally, as in the case of symptom scales, we also sought to categorize participants according to diagnoses. These categories were developed before the current analyses and discussed in a previous paper.[36] The T1000 also included people with eating disorders, but we excluded them from the present study because of small sample sizes. We also categorized mood and anxiety disorders together for our analyses because of the high rates of overlap in these diagnoses and because the sample size for anxiety alone would have been very small (n = 19) if separated. We included a lower number of healthy controls to maximize our ability to detect dimensional effects in patient populations in other planned analyses (in consideration of the total sample size that could be collected).

Data collection procedure

Participants underwent an intensive assessment for demographic, clinical and psychiatric features, with a main focus on negative and positive affect, arousal and cognitive functioning. From this assessment, we acquired several direct and derived variables, only some of which were used in the present analyses. The complete list of assessments and references supporting their validity and reliability are provided in Victor and colleagues.[31]

Approach-avoidance conflict task

The AAC task (Fig. 1)[11,30] is described more extensively in Appendix 1, available at jpn.ca/200032-a1. Before performing the task, participants received detailed instructions (provided in Appendix 2, available at jpn.ca/200032-a2) and completed 4 practice trials to ensure sufficient understanding. In each trial, a runway was shown with a picture of an avatar in a starting position above the runway. Pictures were also shown on each side of the runway, indicating the types of stimuli (i.e., affective image–sound combinations and reward points) that could be presented at the end of the trial depending on participants' choices. Specifically, a sun or a cloud represented potential positive or negative affective stimuli,

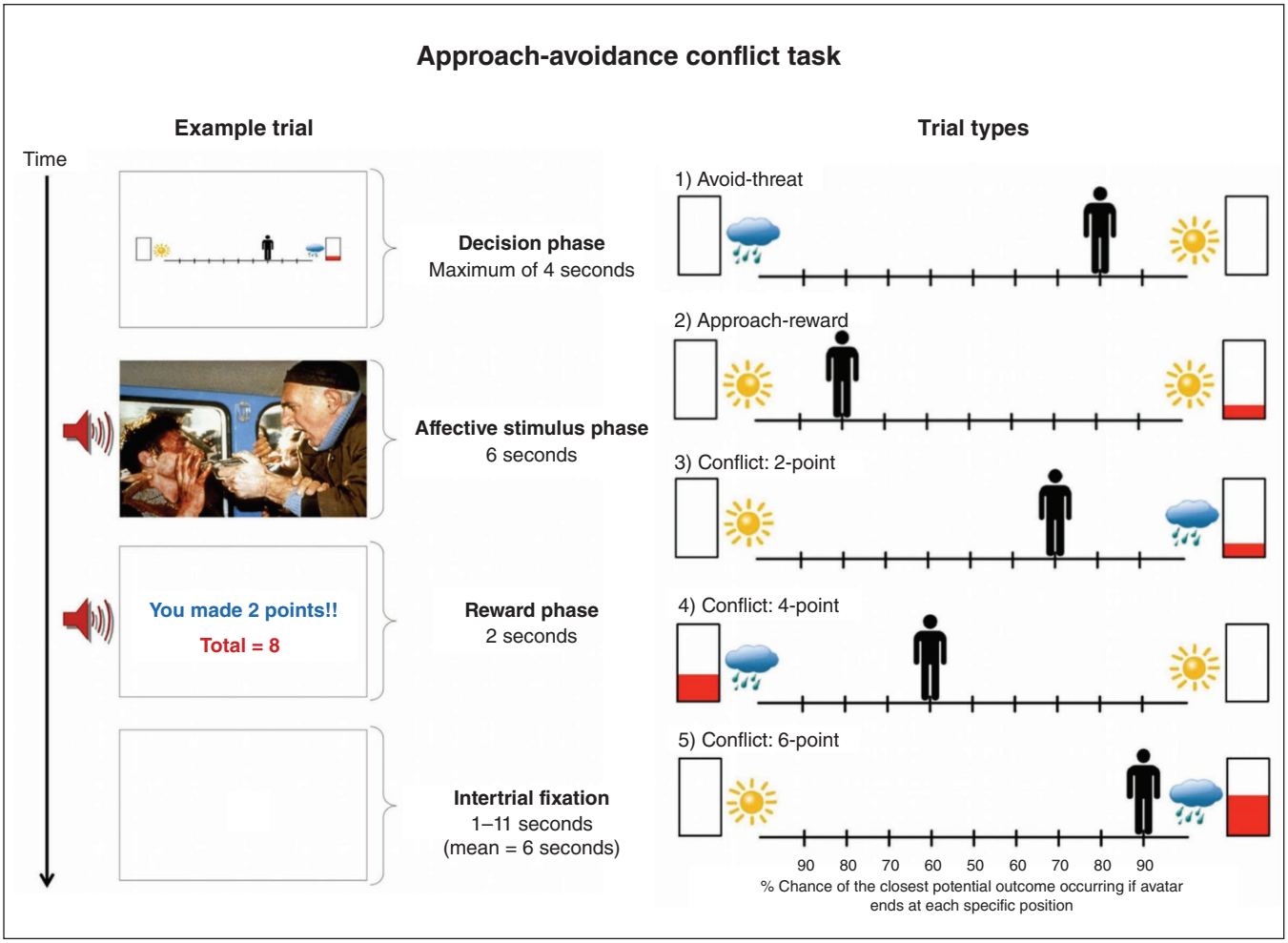

Fig. 1: Left: Sample trial in the approach-avoidance conflict task, in which the negative stimulus and 2 points were presented based on the probabilities associated with the chosen runway position. Right: The 5 trial types. The sun indicates a positive stimulus, the cloud indicates a negative stimulus and the higher the red bar is filled the more points may be received.

respectively, and the height of red fill in a rectangle signified the number of points that would be received in addition to the presentation of those stimuli. In each trial, participants could press the left or right arrow keys to move the avatar from its starting position to any other position (9 possible locations) on the runway, and they were asked to choose a single ending position in each trial. They were told that each ending position corresponded to a specific probability of observing different stimuli at the end of the trial. These stimuli included a positive or negative affective image–sound combination (indicated by the sun or the cloud, respectively), and a certain level of reward points (indicated by the height of the red fill in the rectangle, as noted above). The ending position of the avatar determined the probability that each of these outcomes would occur.

Before they started the task, participants were told the specific probabilities of observing each stimulus for each runway position, and that these probabilities were stable across the task. Thus, there was no learning in this task, and no measure of better or worse performance; participants simply indicated their preferred location on the runway (based on the probabilities of each outcome) in each trial. The probabilities given to the participants were based on the distance from each stimulus (e.g., being closer to the sun image indicated a higher probability of observing the positive stimulus). From left to right on the runway, the probabilities were as follows: 0.9/0.1, 0.8/0.2, 0.7/0.3, 0.6/0.4, 0.5/0.5, 0.4/0.6, 0.3/0.7, 0.2/0.8, 0.1/0.9. The starting position of the avatar (middle, left end or right end) was counterbalanced across trials (for each trial type; see below) to control for its potential influence on participants' choices.

The affective image–sound combinations were gathered from the International Affective Picture System,[48] the International Affective Digitized Sounds[49] and other freely available audio files (see further descriptions in previous reports using this task: Aupperle and colleagues[11] and Chrysikou and colleagues[30]). The "reward" included 0, 2, 4 or 6 points, presented along with a trumpet sound. There were 5 trial types (Fig. 1), which were indicated to participants by the images shown on each side of the runway on each trial. Each trial type was named in reference to the behavioural motivation presumably elicited by the negative or positive affective stimuli and/or the reward points: "avoid-threat" (AV), in which 0 points were offered for both possible stimulus outcomes, so the only explicit motivation was to avoid the negative affective stimulus; "approach-reward"(APP), in which 2 or 0 points were offered, each with positive affective stimuli, so the only explicit motivation was to approach the rewarded outcome; and 3 trial types with different levels of "conflict," in which the negative affective stimulus was presented in addition to receiving either 2 (CONF2), 4 (CONF4) or 6 (CONF6) points (0 points were offered for the other possible outcome, in which a positive affective stimulus would be presented). The task consisted of a total of 60 trials: 12 of each of the 5 trial types.

After task completion, a screen appeared displaying the total points received and an award ribbon. As in previous administrations of the task,[11,12] points did not correspond to a monetary reward. Notably, previous research has shown that paradigms involving non-monetary or monetary rewards elicit similar neural activation patterns in reward-sensitive brain regions,[50,51] which could suggest similar motivational influences. Behavioural variables consisted of the chosen avatar position and the reaction time (i.e., time to initial button press) in each trial. Participants were also asked to fill out a questionnaire that asked about their experiences during the task.

Computational modelling

To model behaviour on the AAC task described above, we adopted a Markov decision process model under the active inference framework; for more details about the structure and mathematics of this class of models, see Friston and colleagues,[23,52] Parr and Friston[53] and Appendix 1. We chose this model because it is well suited for modelling decision-making under uncertainty and was designed to model inference and planning processes, both with and without learning. This was appropriate here because the AAC task did not involve learning (i.e., participants were explicitly told the probabilities of stimuli/points given each runway position before the task began), which made other common learning-based modelling approaches (e.g., reinforcement learning) less appropriate. This was similar to other previously used behavioural tasks (e.g., the urn or beads task[54–56]) that rely on probabilistic inference as opposed to learning, and therefore call for some form of Bayes optimality assumptions (under the complete class theorem; see Huq and colleagues[55]). Furthermore, because the outcomes of decisions in the AAC task were probabilistic and participants were explicitly informed about these probabilities when making their choices, a model that explicitly incorporated action–outcome probabilities appeared to be most appropriate for capturing the cognitive processes that underlie participants' behaviour.

Briefly, this approach required creating a model with specific sets of observable stimuli (o) and beliefs about states of the task (s), as well as beliefs about the sequences of actions that can be chosen (policies; π). For this study, observations included runway position, trial-type cues and positive/negative stimuli + number of points; task states included beliefs about the trial type and position on the runway; and policies included transitions to each possible location on the runway. The relationships between these variables at a time (t) were described by a set of matrices. The **A** matrix encoded the way that task states were related to observations, $P(o_t | s_t)$. This included (1) the probability of observing each set of trial-type cues given a particular trial type (specified as an identity matrix); (2) the probability of observing the avatar in a given location given a particular runway position (specified as an identity matrix); and (3) the probability of observing different outcome stimuli given each position on the runway (specified based on the stated task probabilities). The **B** matrix encoded the probability that one task state would transition into another depending on selected policies, $P(s_{t+1} | s_t, \pi)$; in this study, this specified that trial type was stable across a trial (identity matrix) and that the participant would deterministically transition from the start state to different runway

position states given the selection of different policies. The value of an observation was encoded as a log probability within a matrix referred to as the **C** matrix (implementation in the present model described further below). Policies were selected based on beliefs about the probability that each possible policy would produce preferred observations (i.e., formally, those with the highest prior probability), modulated by an expected precision term (γ) that encoded decision uncertainty. See Table 1 and Appendix 1 for further details about how observations, states, policies and associated matrices were defined to model the AAC task.

Each trial consisted of 2 epochs. In the first epoch, the participant was in a "start" state and was presented with the avatar and trial-type cues (observations indicating AV, APP, or CONF2, 4 or 6). Based on the selected policy, on the second epoch the participant then transitioned to the chosen runway position (indicated by movement of the avatar) and observed the outcome stimuli (i.e., image–sound type + number of points) based on the probabilities associated with the chosen runway position. Figure 2 depicts the Markov decision process structure, **A**-matrices for the task and sample simulations under different parameter values. Figure 3 provides a visual depiction of the model structure of the AAC task.

The **C** matrix was specified such that the expected value assigned to each possible stimulus was determined by 3 parameters corresponding to the subjective value of observing the positive affective stimulus, the negative affective stimulus and each point that could be won. The positive affective stimulus was fixed at an "anchor" value of $\ln P(o) = 0$, and the value of each point was $\ln P(o) = 1$ (e.g., $\ln P[o] = 2$ when winning 2 points). We then estimated the relative value (subjective aversiveness) of the negative affective stimulus. This parameter indicated the "emotional conflict" (EC) — that is, the expected aversiveness of the affective stimuli relative to the reward value assigned to each point. We also estimated a prior policy precision parameter (β), which was the inverse of the expected precision term γ and acted as an index of an individual's a priori belief-based uncertainty related to making optimal decisions. Higher decision uncertainty in the model (a higher β value) led to less consistent (more variable) choices over trials, because of less precise beliefs about the best policy (i.e., less confidence in which action would lead to the most preferred outcomes).

Based on these values, the model inferred a probability distribution over possible actions (i.e., transitions to different possible runway positions) and sampled actions from this distribution on each trial, where higher action probabilities formally corresponded to lower values of a quantity called expected free energy (G; described more thoroughly in Appendix 1). Briefly, in this context a lower expected free energy corresponded to a smaller (Kullback–Leibler) divergence between the distribution reflecting the preferred observations

Table 1: Markov decision process model of the approach-avoidance conflict task

Model variable	General definition	Model-specific specification
o_t	Observable outcomes at time t	Outcome modalities: 1. Observed position on the runway (10 possible observations, including a "starting" position and the 9 final positions one could choose) 2. Cues indicating trial type (5 possible observations, corresponding to the 5 trial types) 3. Stimuli observed at the end of each trial. This included 7 possible observations corresponding to a "starting" observation, the positive stimulus with 0 or 2 points, and the negative affective stimulus with 0, 2, 4 or 6 points
s_t	Hidden states at time t	Hidden state factors: 1. Beliefs about position on the runway (10 possible belief states with an identity mapping to the observations in outcome modality 1) 2. Beliefs about the trial type (corresponding to the 5 trial types)
π	A distribution over action policies encoding the expectation that a particular policy is most likely to generate preferred outcomes	Allowable policies included the decision to transition from the starting state to each of the 9 possible positions on the runway
β	The prior on expected policy precision (β) is the "rate" parameter of a γ distribution, which is a standard distribution to use as a prior for expected precision. This latter term modulates the influence of expected free energy on policy selection	When β is high (reflecting low confidence about the best decision), policy selection becomes less deteriministic. Higher β values therefore encode participants' decision uncertainty during the task (similar to the temperature parameter in a conventional softmax response function)
A matrix $P(o_t \mid s_t)$	A matrix encoding beliefs about the relationship between hidden states and observable outcomes (i.e., the likelihood that specific outcomes will be observed given specific hidden states)	Encodes beliefs about the relationship between position on the runway and the probability of observing each outcome, conditional on beliefs about the task condition
B matrix $P(s_{t+1} \mid s_t)$	A matrix encoding beliefs about how hidden states will evolve over time (transition probabilities)	Encodes beliefs about the way participants could choose to move the avatar, as well as the belief that the task condition will not change within a trial
C matrix $\ln P(o_t)$	A matrix encoding the degree to which some observed outcomes are preferred over others (technically modelled as prior expectations over outcomes)	Encodes stronger positive preferences for receiving higher numbers of points, and negative preferences for the aversive stimuli (both relative to an anchor value of 0 for the "safe" positive stimulus). The emotional conflict (EC) parameter in our model encoded the value of participants' preferences against observing the aversive stimuli
D matrix $P(s_1)$	A matrix encoding beliefs about (a probability distribution over) initial hidden states	The simulated agent always began in an initial starting state, and believed each task condition was stable across each trial

(i.e., the combined negative + positive values of images, sounds and points) and the distribution reflecting the observations expected under each choice of runway position (given knowledge of the associated probabilities). Lower expected free energy of an action thus indicated a higher probability of observing the overall most preferred combination of stimuli/points if that action were chosen.

Our computational phenotyping approach used Bayesian inference at 2 levels.[58] First, each participant's responses were modelled under ideal Bayesian assumptions, using the

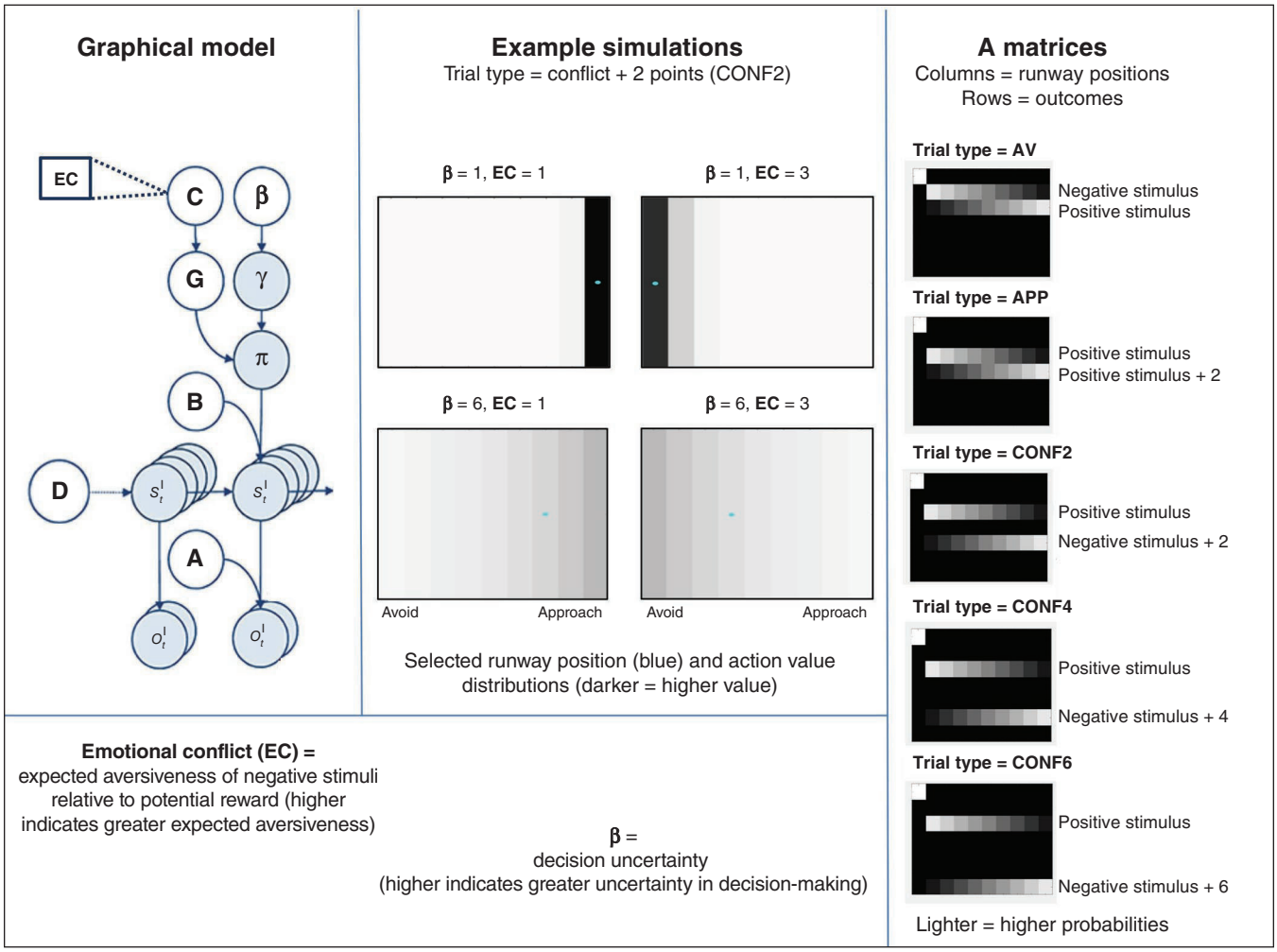

Fig. 2: Computational model. Top left: The Markov decision process used to model the approach-avoidance conflict task. The generative model is depicted graphically, such that arrows indicate dependencies between variables. Observations (*o*) depend on hidden states (*s*; this relationship is specified by the **A** matrix), and those states depend on both previous states (as specified by the **B** matrix or the initial states specified by the **D** matrix) and the sequences of actions/policies (π) selected by the agent. The probability of selecting a particular policy in turn depends on the expected free energy (*G*) of each policy with respect to the preferences (**C**) of the decision-maker being modelled. The degree to which expected free energy influences policy selection is also modulated by an expected precision term (γ), which is in turn dependent on a prior policy precision parameter (β), where higher values of β promote greater decision uncertainty (i.e., less influence of the differences in expected free energy across policies). For more details on the associated mathematics, see Friston and colleagues[52,57] and Appendix 1, available at jpn.ca/200032-a1. In our model, the observations were cues indicating the trial type, cues indicating the position of the avatar, and the outcome stimuli. The hidden states included beliefs about trial type and avatar position, and the policies included the choice to move the avatar to any other position on the runway. Right: The **A** matrices in the right panel show the mapping between states and observations for outcome stimuli. Here, the rows correspond to the stimuli (first row is the "start" observation), and the columns correspond to the avatar position states (column 1 corresponds to the "start" state, and columns 2 to 10 correspond to choosing each of the 9 runway positions). Lighter colours in these **A** matrices indicate higher probabilities. Trial types: AV = avoid; APP = approach; CONF2, CONF4 and CONF6 = conflict + 2, 4 or 6 points, respectively. Bottom left: The model parameters corresponded to the degree to which the negative stimulus was dyspreferred relative to the degree to which the points were preferred in the **C** matrix ("emotional conflict"; EC), as well as the prior policy precision parameter β, which reflected decision uncertainty. Top middle: Example simulations of action selection under different parameter values during the CONF2 trial type. Blue dots indicate chosen actions, and darker colours indicate higher action values in the model.

Markov decision process formulation of choice behaviour described previously. Based on the trial-by-trial task stimuli observed by each participant and their trial-by-trial decisions, we then used variational Bayes to estimate each participant's prior beliefs that maximized the likelihood of their responses, as described in Schwartenbeck and Friston.[18] In other words, the observation model for estimating subject-specific preferences and precision was based on the assumption that subjects were using (active) Bayesian inference. In this setting, active inference can be seen as a generalization of Bayesian decision theory that replaces the expected value or utility with expected log evidence or marginal likelihood for a generative model of the task.[18,59] This means that subjective responses are sampled from posterior beliefs about the best course of action, where these posterior beliefs depend on their prior preferences about the consequences of a decision and the information gain afforded by their actions. This posterior distribution over behavioural responses can then be used to assess the likelihood of responses under different prior beliefs. We optimized these preferences (and precision of posterior beliefs about policies) using this likelihood and standard variational Laplace.[60] Having estimated each participant's preferences (and precision), we then used classical inference to test for the effects of group, using a standard summary statistics-based approach.

As described in more detail in Appendix 1, we also considered 2 other models: a simpler 1-parameter model including no decision uncertainty term (estimating EC only), and a more complex 3-parameter model that fit separate terms for the subjective value of the negative affective stimuli and the subjective value of the points (i.e., in place of the EC parameter). Initial simulations indicated that parameter estimates for the 3-parameter model were dependent on prior values (and therefore not recoverable), because only the relative

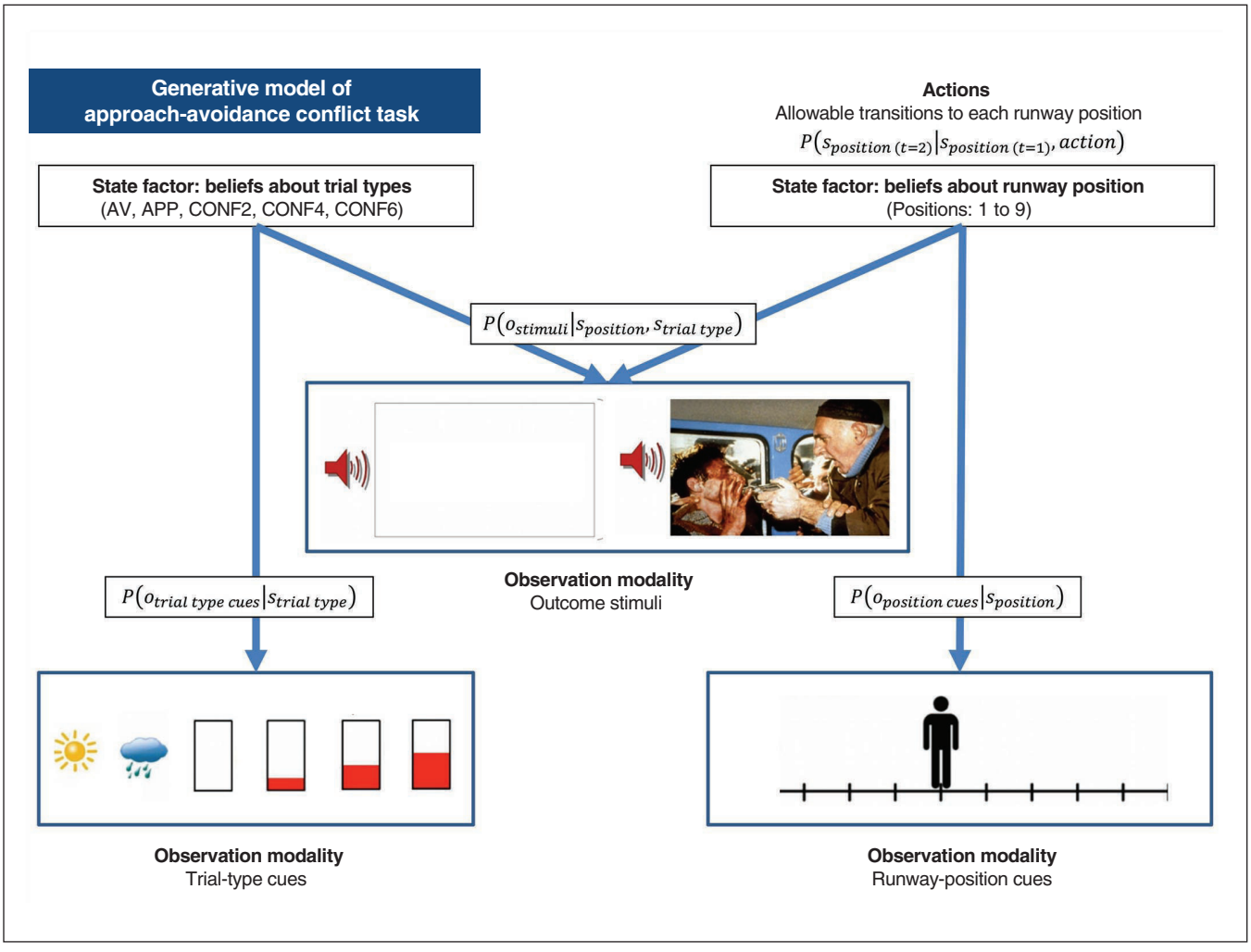

Fig. 3: Simplified visual depiction of relevant dependencies in the computational (generative) model of the approach-avoidance conflict task. Beliefs about trial type and beliefs about runway positions were generated by (and inferred based on) trial-type cues and runway-position cues, respectively. Observed outcome stimuli were probabilistically generated by an interaction between trial type and runway position. Beliefs about this interaction were used to infer the action (state transition) most likely to produce the most preferred outcome stimuli. Trial types: AV = avoid; APP = approach; CONF2, CONF4 and CONF6 = conflict + 2, 4 or 6 points, respectively.

value of the negative stimuli versus points ultimately influenced behaviour. Therefore, we did not use this model. In contrast, estimates from the simpler 1-parameter model did appear recoverable. However, Bayesian model comparison (based on Rigoux and colleagues[61] and Stephan and colleagues[62]) showed that this model performed worse than the 2-parameter model (protected exceedance probability = 1).

We implemented all model simulations using standard routines (spm_MDP_VB_X.m) that are available as Matlab code in the latest version of SPM academic software (www.fil.ion.ucl.ac.uk/spm/). Matlab code specifying the generative model of the AAC task is included in Appendix 3, available at jpn.ca/200032-a3 (AAC_model.m).

Statistical analysis

We conducted statistical analyses using the R statistical package (2018; www.R-project.org/). To assess face validity, we calculated a model accuracy score that reflected the average percentage of trials during which the action with the highest probability in the model matched the action chosen by participants (i.e., under the parameter values estimated for each participant) and we examined correlations between model parameters and reaction times (i.e., time to initial button press, both across the whole task and within each condition) with the expectation that valid computational measures of greater emotional conflict (EC) and decision uncertainty (β) would both be associated with slower reaction times. We then conducted further correlation analyses to examine whether each parameter could predict subsequent self-reports on the post-task Likert scale questions. The validity of EC would be supported by positive associations with self-reported avoidance motivation and anxiety, and the validity of β would be supported by self-reported difficulty making decisions and self-reported avoidance motivation.

We used R statistical software to conduct analyses of covariance to identify possible group differences in each parameter while accounting for individual differences in age, sex, the Wide Range Achievement Test reading score (WRAT; a common measure of premorbid IQ[63]) and the interaction between group and each of these factors. We used WRAT scores to ensure that differences in task behaviour could not be accounted for by differences in general intelligence. Results were Bonferroni-corrected for multiple comparisons with 2 parameters ($\alpha = 0.05$, $p < 0.025$). To assess group differences in approach-avoidance behaviour, we also conducted similar analyses of covariance with standard descriptive task variables as the dependent variables, including mean chosen runway positions during the AV, APP and CONF trial types (higher values indicated stronger approach behaviour toward the points; in the AV condition, higher values indicated positions closer to the positive stimulus).

The groups showed significant differences in age, sex and WRAT scores, which prevented strong conclusions when trying to control for the effects of these variables.[64] This confirmed the group differences anticipated based on the representative demographics of these clinical populations.[65–72] As has been done in previous work on the T1000 data set,[73] to more rigorously assess group differences we used the fullmatch function in the optmatch R package (www.rdocumentation.org/packages/optmatch/versions/0.9-10/topics/fullmatch) to propensity-match groups based on age and WRAT scores (propensity-matching was not effective when including sex, given the differences between groups). Propensity-matching led to sample sizes of 59 healthy controls, 161 participants with depression and/or anxiety and 56 participants with substance use disorders) for the matched samples. We then performed the analyses described above with the propensity-matched groups, with sex and group × sex interaction as independent variables in the model.

The T1000 study — of which the AAC task analyzed here is a part — was designed explicitly in an exploratory/confirmatory framework; the first 500 participants were designated as exploratory, and the second 500 participants were reserved for confirmatory analyses based on the results of the exploratory analyses. As such, the analyses reported here should be considered exploratory. Confirmatory analyses will be carried out in planned future work.

Results

Descriptive statistics for demographic and clinical measures are shown in Table 2. The descriptive statistics for each of the parameters were as follows (mean ± standard deviation): β = 4.77 ± 4.69; EC = 2.70 ± 2.72. For further information about the relationship between model parameters and demographic variables, see Appendix 1. The EC and β parameters were correlated at $r = 0.26$ and $p < 0.001$. Because the parameters were not normally distributed, they were log-transformed for all subsequent analyses using the R package optLog (https://github.com/kforthman/optLog) to find the optimal log-transform that minimized skew. This package was originally developed by researchers at the Laureate Institute for Brain Research. Parameter distributions before and after transformation are shown in Appendix 1.

Face validity: task-related self-report and behaviour

Averaging across participants, the model was accurate at predicting behaviour in 72% (standard error = 1%) of trials (note: chance accuracy is 1/9 = 11%). Participants with longer reaction times across all trials also exhibited greater EC ($r = 0.24$, $p < 0.001$) and higher β values ($r = 0.59$, $p < 0.001$). Analyses of reaction times within specific trial types showed similar results (see Appendix 1). Relationships between model parameters and self-reports on the post-AAC questionnaire items are shown in Table 3. Notably, EC correlated most strongly with self-reported motivations to move toward reward ($r = -0.74$, $p < 0.001$) and away from negative outcomes ($r = 0.67$, $p < 0.001$). Higher EC also corresponded to higher self-reported anxiety during the task ($r = 0.32$, $p < 0.001$); β correlated most strongly with self-reported difficulty making decisions on the task ($r = 0.45$, $p < 0.001$) and (reduced) motivations to move toward reward ($r = -0.48$, $p < 0.001$).

Clinical validity: diagnostic effects

Group difference results for the propensity-matched sample are shown in Figure 4. For analogous results in the full sample, see Appendix 1. Results in the full sample showed a highly similar pattern, as we note more specifically below.

We found a main effect of group on β ($F_{2,270} = 4.15$, $p = 0.017$), reflecting lower values in healthy controls than in those with depression and/or anxiety ($t_{128} = 2.16$, $p = 0.03$, $d = 0.30$) or substance use disorders ($t_{102} = 2.88$, $p = 0.005$, $d = 0.53$); effects of sex and the group × sex interaction were nonsignificant. We observed a similar pattern in the full sample (Appendix 1).

We found a main effect of sex on EC ($F_{1,270} = 11.17$, $p < 0.001$; higher in females), but the effects of group and the group × sex interaction were nonsignificant. We did find a trend effect of group in the full sample ($F_{2,466} = 3.30$, $p = 0.04$), reflecting greater EC in healthy controls than in those with substance use disorders ($t_{92} = 3.17$, $p = 0.002$, $d = 0.51$; see Appendix 1 for full results).

In Appendix 1, we have also presented these analyses separately for males and females; the pattern of findings for the EC parameter remained significant only in females, and the pattern of findings for the β parameter remained significant only in males.

Standard descriptive analyses

Descriptive statistics for task-related self-report and traditional performance variables (reaction time, approach behaviour) are provided in Appendix 1. Average reaction times

Table 2: Summary statistics and group differences for demographic and clinical measures

Sample	Healthy controls	Depression/anxiety	Substance use disorders	*p* value
Full sample	*n* = 59	*n* = 260	*n* = 159	
Age, yr	32.14 ± 11.13	35.89 ± 11.30	33.93 ± 9.09	0.024
Male, *n* (%)	28 (48)	70 (27)	74 (47)	< 0.001
PHQ score	0.90 ± 1.36	12.63 ± 5.14	6.50 ± 5.66	< 0.001
OASIS score	1.27 ± 1.88	9.80 ± 3.42	5.78 ± 4.66	< 0.001
DAST-10 score	0.12 ± 0.38	0.67 ± 1.41	7.48 ± 2.20	< 0.001
WRAT score	62.37 ± 5.06	63.53 ± 4.76	58.49 ± 5.65	< 0.001
Propensity-matched sample	*n* = 59	*n* = 161	*n* = 56	
Age, yr	32.14 ± 11.13	35.11 ± 10.84	32.67 ± 10.26	0.12
Male, *n* (%)	28 (48)	41 (25)	35 (63)	< 0.001
PHQ score	0.90 ± 1.36	12.64 ± 5.38	7.95 ± 6.50	< 0.001
OASIS score	1.27 ± 1.88	9.78 ± 3.42	6.80 ± 5.15	< 0.001
DAST-10 score	0.12 ± 0.38	0.62 ± 1.26	7.45 ± 2.65	< 0.001
WRAT score	63.53 ± 4.76	62.58 ± 4.53	61.89 ± 4.43	0.15

DAST-10 = Drug Abuse Screening Test; OASIS = Overall Anxiety Severity and Impairment Scale; PHQ = Patient Health Questionnaire; WRAT = Wide Range Achievement Test.
Unless otherwise indicated, findings are mean ± standard deviation.

Table 3: Post-task self-report questionnaire items and Pearson correlations with the computational model parameters*

		Pearson correlations	
Question†	Mean ± SD	Emotional conflict (EC)	Decision uncertainty (β)
1. I found the POSITIVE pictures enjoyable	5.05 ± 1.68	0.07	0.02
2. The NEGATIVE pictures made me feel anxious or uncomfortable	4.43 ± 1.99	0.32‡	0.06
3. I often found it difficult to decide which outcome I wanted	2.51 ± 1.73	0.10§	0.45‡
4. I always tried to move ALL THE WAY TOWARD the outcome with the LARGEST REWARD POINTS	4.76 ± 2.30	–0.74‡	–0.48‡
5. I always tried to move ALL THE WAY AWAY FROM the outcome with the NEGATIVE PICTURE/SOUNDS	2.98 ± 2.17	0.67‡	0.37‡
6. When a NEGATIVE picture and sound were displayed, I kept my eyes open and looked at the picture	5.5 ± 1.83	–0.37‡	–0.17‡
7. When a NEGATIVE picture and sound were displayed, I tried to think about something unrelated to the picture to distract myself	2.96 ± 1.94	0.29‡	0.11§
8. When a NEGATIVE picture and sound were displayed, I tried other strategies to manage emotions triggered by the pictures	3.26 ± 1.99	0.32‡	0.05

SD = standard deviation.
*Full sample; *n* = 478.
†Answers provided on a Likert scale (1 = not at all; 7 = very much).
‡*p* < 0.001.
§*p* < 0.05.

across trial types were not significantly different between groups; we found no effect of sex or group × sex interaction. However, in Appendix 1 we present some significant group differences in specific trial types, reflecting slower reaction times in the clinical groups relative to the healthy controls. We observed similar results in the full sample (Appendix 1).

Within conflict trials (CONF2, CONF4, CONF6), we observed a main effect of sex on chosen runway position ($F_{1, 270}$ = 7.53, p = 0.006; less avoidance in males). Within AV trials, we observed a main effect of sex ($F_{1, 270}$ = 5.29, p = 0.02; less avoidance of the negative image in males) and group ($F_{2, 270}$ = 13.36, p < 0.001) on chosen runway position. Further inspection indicated that healthy controls showed greater avoidance of the negative image in this condition than those with depression and/or anxiety (t_{177} = 2.42, p = 0.016, d = 0.29) or those with substance use disorders (t_{72} = 4.90, p < 0.001, d = 0.93). Those with substance use disorders also showed less avoidance than those with depression and/or anxiety (t_{75} = 3.59, p < 0.001, d = 0.65). Within APP trials, we observed a main effect of group on chosen runway position ($F_{2,270}$ = 5.04, p = 0.007), reflecting greater approach behaviour (i.e., toward the points) in healthy controls than in those with depression and/or anxiety (t_{203} = 3.41, p < 0.001, d = 0.38) and those with substance use disorders (t_{81} = 3.67, p < 0.001, d = 0.69). We found somewhat similar results in the full sample (Appendix 1).

Although decision uncertainty was not analyzed in previous studies of this AAC task, our results for this parameter suggested the presence of differences in within-subject decision variability. See Appendix 1 for additional analyses that confirm these differences (most strongly in the AV, APP and CONF6 trial types) and also show the expected positive correlations between within-subject variability and β values across all trial types.

Discussion

Using a novel active inference modelling approach in a large community sample, this study uncovered separable influences of 2 factors on approach-avoidance behaviour: expected outcome aversiveness relative to expected reward (EC) and decision uncertainty (β). The model showed high accuracy in predicting behaviour (72%; i.e., relative to chance accuracy = 11%) and was further validated by the fact that parameter estimates showed strong relationships with reaction times,

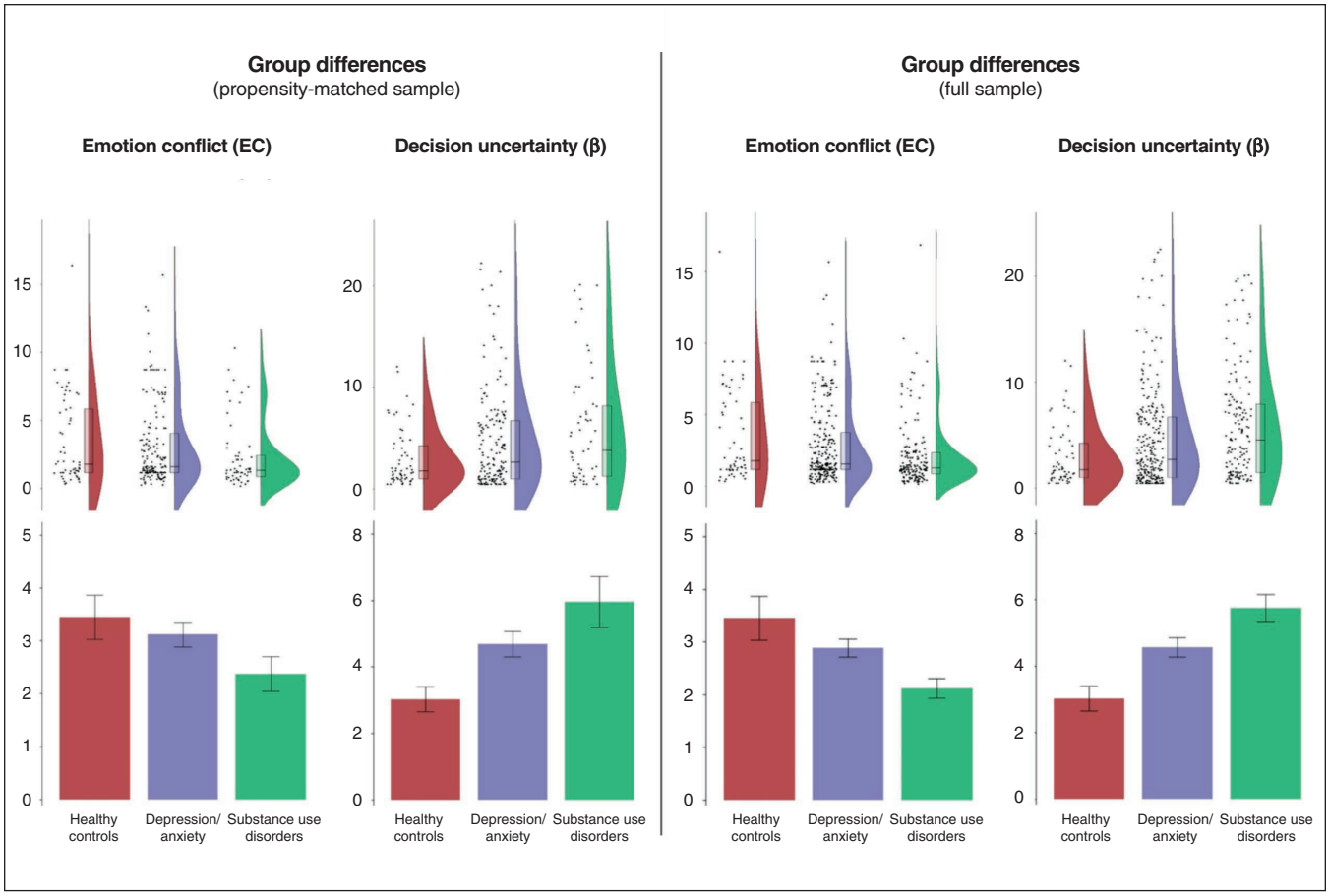

Fig. 4: Raincloud plots (distributions, box plots and individual data points) and bar plots (means and standard errors) showing differences between healthy controls and clinical groups in emotional conflict (EC; expected aversiveness of negative stimuli relative to reward) and decision uncertainty (β; expected policy precision) in the active inference model of the approach-avoidance conflict task. Left: Propensity-matched sample. Right: Full sample. Data displayed are before log-transformation.

participants' self-reported feelings/motivations during the task and self-reported approach-avoidance motivations, all in the expected directions. Further, EC was uniquely associated with self-reported anxiety on the task, while β was uniquely associated with self-reported difficulty making decisions on the task. Crucially, EC and β were not highly correlated, with distinct relationships to psychopathology.

The computational approach used here may have provided advantages over previous descriptive analyses of task behaviour, because it could disentangle the effects of conflict and uncertainty.[11,30] Although previous analyses have focused mainly on relative approach-avoidance drives (i.e., captured by chosen runway position) and reaction times, our approach uncovered a unique pattern of differences in decision uncertainty that may have been missed in previous studies. Specifically, our results showed that patients with depression, anxiety and substance use disorders exhibited greater uncertainty in decision-making relative to healthy controls, and that those with substance use disorders tended toward lower emotional conflict (although this latter finding was marginal). Standard descriptive analyses of reaction times and chosen runway position did not pick up on this difference, and, contrary to expectation, did not show evidence of greater avoidance in the clinical groups. Instead, the clinical groups showed no behavioural differences on conflict trials and showed less of the expected approach and avoidance drives in the 2 non-conflict conditions. Post-hoc analyses of within-subject choice variability that were motivated by our decision uncertainty results (which has not been examined in previous studies) also showed greater choice variability in the clinical groups in these non-conflict conditions (as well as in the CONF6 condition; see Appendix 1), suggesting that a mix of separable approach and avoidance drive abnormalities may contribute to maladaptive behaviour during AAC. Thus, our model-based findings uncovered a pattern of decision uncertainty that might be more clinically relevant to avoidance behaviour in real-world settings.

The finding that both clinical groups showed greater β values relative to healthy controls suggests reduced confidence in their internal model of how to act, and a resulting inconsistency in choice behaviour. This is because, formally, the β parameter in active inference models reflects prior expectations about one's ability to select the best action. This is consistent with the correlation we observed between slower reaction times and higher β values. Thus, at least in some situations, maladaptive approach-avoidance behaviour in substance use disorders, depression and anxiety could relate more to decision uncertainty than to increased emotional conflict (e.g., avoidance motivation, threat sensitivity) per se. It is also worth noting that our supplementary analyses suggested this effect was stronger in males, which could relate to previous work suggesting that higher anxiety sensitivity more strongly drives avoidance behaviour in males than females,[12] a finding that could have important treatment implications (e.g., that different interventions may be necessary in males and females).

It is useful to consider whether current or novel interventions aimed at modifying decision uncertainty may have potential for improving mental health. Many current interventions for depression and/or anxiety focus on either threat (i.e., exposure-based therapy) or reward reactivity (i.e., behavioural activation approaches), but it is possible that these or other treatment strategies (e.g., cognitive restructuring or problem-solving) may in fact target decision uncertainty. Investigations into the relationships between decision uncertainty and response to these treatments is important for further delineating clinical implications. However, the source of the greater uncertainty in the clinical groups is unclear. Based on previous work,[42] one possibility could be that psychopathology involves difficulty in separating "signal from noise" when observing the outcomes of actions (i.e., did the outcome of my previous action come about by chance, or would the same outcome happen again?). However, this interpretation (among other possible interpretations) will need to be tested in future work.

Contrary to our initial hypothesis, the model did not offer evidence that the clinical groups had higher EC. In contrast, we found suggestive evidence for lower emotional conflict in substance use disorders than in healthy controls. Because this finding occurred only in the full sample, we do not offer strong interpretations here. However, we briefly note that it may be consistent with previous studies suggesting a general blunting of brain and behavioural responses to affective stimuli in people with cocaine and methamphetamine use disorders,[74,75] and with studies that have linked lower self-reported sensitivity to punishment with both methamphetamine and marijuana use,[76,77] which could relate to continued drug use because of an insensitivity to its negative consequences (for recent computational modelling evidence supporting this possibility, see Smith and colleagues[78]). This finding could also relate to previous work demonstrating that cognitive bias modification methods that train increased avoidance in response to alcohol cues is beneficial to recovery in an inpatient sample of people with alcoholism.[79] Interestingly, our supplementary analyses suggested that this result was stronger in females, which bore some similarity to previous results suggesting that reduced reward motivation plays a larger role in avoidance behaviour in females.[12]

Our results may also have implications for the active inference literature. For example, previous proposals have suggested a link between state anxiety and the β parameter (or related computational parameters associated with emotion and uncertainty),[80–86] where higher uncertainty is suggested to underpin higher anxiety. However, we found that higher EC, but not higher β, was associated with higher self-reported anxiety during the task (on the post-task Likert scale questionnaire). Our results therefore linked state anxiety to the preference distribution in the model, suggesting that some people with strong anxiety responses may have found the stimuli quite aversive, and yet were highly confident and consistent in their avoidant strategy. Future empirical work in active inference research should therefore disentangle the circumstances in which negative affect is associated with uncertainty versus avoidance drives in these models. A second implication for active inference stems from previous work linking fluctuations (updates) in β values to phasic dopamine responses.[87–90] Because substance use has been linked to dopaminergic system dysfunction (e.g., Huys and colleagues[91] and Koob and Volkow[92]), this model/task could be used during neuroimaging to simulate predicted individual

differences in dopaminergic dynamics (associated with differences in β[90]) and perhaps shed further light on the potential role of dopamine and its relation to decision uncertainty in contributing to symptoms and/or treatment response (e.g., the potential role of dopaminergic medications in altering decision uncertainty).

Limitations

This was the first study to integrate an active inference model with an affect-based AAC paradigm in a large, transdiagnostic, community-based clinical sample. Although the validity of the model parameters was supported by the model's high accuracy in predicting behaviour and by the parameters' expected relationships to standard task measures, there are important limitations to consider. As is inevitable when performing model-based analyses, we were required to make certain choices about fixed parameter values. We also chose to use a Bayesian modelling approach because of the probabilistic nature of the AAC task, but other modelling approaches could have been considered. Still, our model comparison results and the correlations observed between model parameters and other self-report/behavioural measures suggest that these choices were reasonable. Second, the AAC task used in this study was not designed in advance with modelling explicitly in mind. Future work should investigate potential task modifications that could further disentangle distinct computational influences on approach-avoidance behaviour. Because this study was exploratory, our results will also need to be replicated by other researchers, as well as in our planned confirmatory analyses on participants sampled from the second 500 participants of the T1000 data set.

Conclusion

Our results demonstrate a novel method of modelling affect-based AAC behaviour that was able to differentiate distinct components of conflict. Relative to healthy controls, transdiagnostic behavioural differences during AAC were better accounted for by greater decision uncertainty, as opposed to greater sensitivity to negative affective stimuli. Future research should replicate these findings and further investigate their potential clinical relevance.

Affiliations: From the Laureate Institute for Brain Research, Tulsa, OK, USA (Smith, Kirlic, Stewart, Touthang, Kuplicki, Khalsa, Feinstein, Paulus, Aupperle); and the Oxley College of Health Sciences, University of Tulsa, Tulsa, OK, USA (Stewart, Khalsa, Paulus, Aupperle).

Competing interests: None declared.

Funding: This work was funded by the NIGMS (P20 GM121312; PI:MPP), the NIMH (K23-MH108707; PI: RLA), and the William K. Warren Foundation.

Contributors: R. Smith, R. Kuplicki, J. Feinstein, M. Paulus and R. Aupperle designed the study. J. Touthang, R. Kuplicki and M. Paulus acquired the data, which R. Smith, N. Kirlic, J. Stewart, J. Touthang, R. Kuplicki, S. Khalsa, M. Paulus and R. Aupperle analyzed. R. Smith, N. Kirlic, J. Stewart, M. Paulus and R. Aupperle wrote the article, which all authors reviewed. All authors approved the final version to be published and can certify that no other individuals not listed as authors have made substantial contributions to the paper.

References

1. Aupperle RL, Paulus M. Neural systems underlying approach and avoidance in anxiety disorders. *Dialogues Clin Neurosci* 2010;12:517.
2. Barlow DH, Allen L, Choate M. Toward a unified treatment for emotional disorders—republished article. *Behav Ther* 2016;47:838-53.
3. Ekhtiari H, Victor TA, Paulus MP. Aberrant decision-making and drug addiction—how strong is the evidence? *Curr Opin Behav Sci* 2017;13:25-33.
4. Guttman Z, Moeller SJ, London ED. Neural underpinnings of maladaptive decision-making in addictions. *Pharmacol Biochem Behav* 2018;164:84-98.
5. Paulus MP. Evidence-based pragmatic psychiatry—a call to action. *JAMA Psychiatry* 2017;74:1185-6.
6. Kirlic N, Young J, Aupperle RL. Animal to human translational paradigms relevant for approach avoidance conflict decision making. *Behav Res Ther* 2017;96:14-29.
7. Lejuez CW, Read J, Kahler C, et al. Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *J Exp Psychol Appl* 2002;8:75-84.
8. Bechara A, Damasio H, Tranel D, et al. Deciding advantageously before knowing the advantageous strategy. *Science* 1997;275:1293-5.
9. Talmi D, Dayan P, Kiebel SJ, et al. How humans integrate the prospects of pain and reward during choice. *J Neurosci* 2009;29:14617-26.
10. Schlund MW, Brewer AT, Magee SK, et al. The tipping point: value differences and parallel dorsal-ventral frontal circuits gating human approach-avoidance behavior. *Neuroimage* 2016;136:94-105.
11. Aupperle RL, Melrose AJ, Francisco A, et al. Neural substrates of approach-avoidance conflict decision-making. *Hum Brain Mapp* 2015;36:449-62.
12. Aupperle RL, Sullivan S, Melrose AJ, et al. A reverse translational approach to quantify approach-avoidance conflict in humans. *Behav Brain Res* 2011;225:455-63.
13. Rinck M, Becker ES. Approach and avoidance in fear of spiders. *J Behav Ther Exp Psychiatry* 2007;38:105-20.
14. Friston KJ, Stephan K, Montague R, et al. Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry* 2014;1:148-58.
15. Huys QJ, Maia T, Frank M. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci* 2016;19:404-13.
16. Montague PR, Dolan R, Friston K, et al. Computational psychiatry. *Trends Cogn Sci* 2012;16:72-80.
17. Petzschner FH, Weber L, Gard T, et al. Computational psychosomatics and computational psychiatry: toward a joint framework for differential diagnosis. *Biol Psychiatry* 2017;82:421-30.
18. Schwartenbeck P, Friston K. Computational phenotyping in psychiatry: a worked example. *eNeuro* 2016;3:ENEURO.0049-16.2016.
19. Krypotos AM, Beckers T, Kindt M, et al. A Bayesian hierarchical diffusion model decomposition of performance in approach-avoidance tasks. *Cogn Emot* 2015;29:1424-44.
20. Sheynin J, Moustafa AA, Beck KD, et al. Testing the role of reward and punishment sensitivity in avoidance behavior: a computational modeling approach. *Behav Brain Res* 2015;283:121-38.
21. Kirlic N, Aupperle RL, Rhudy JL, et al. Latent variable analysis of negative affect and its contributions to neural responses during shock anticipation. *Neuropsychopharmacology* 2019;44:695-702.
22. Friston K, FitzGerald T, Rigoli F, et al. Active inference and learning. *Neurosci Biobehav Rev* 2016;68:862-79.

23. Friston K, FitzGerald T, Rigoli F, et al. Active inference: a process theory. *Neural Comput* 2017;29:1-49.
24. Amir N, Foa EB, Coles ME. Automatic activation and strategic avoidance of threat-relevant information in social phobia. *J Abnorm Psychol* 1998;107:285-90.
25. Dugas MJ, Gagnon F, Ladouceur R, et al. Generalized anxiety disorder: a preliminary test of a conceptual model. *Behav Res Ther* 1998;36:215-26.
26. Elliott R, Sahakian BJ, McKay AP, et al. Neuropsychological impairments in unipolar depression: the influence of perceived failure on subsequent performance. *Psychol Med* 1996;26:975-89.
27. Paulus MP, Yu AJ. Emotion and decision-making: affect-driven belief systems in anxiety and depression. *Trends Cogn Sci* 2012;16: 476-83.
28. Garami J, Haber P, Myers CE, et al. Intolerance of uncertainty in opioid dependency—relationship with trait anxiety and impulsivity. *PLoS One* 2017;12:e0181955.
29. Wei Z, Han L, Zhong X, et al. Chronic nicotine exposure impairs uncertainty modulation on reinforcement learning in anterior cingulate cortex and serotonin system. *Neuroimage* 2018;169:323-33.
30. Chrysikou EG, Gorey C, Aupperle RL. Anodal transcranial direct current stimulation over right dorsolateral prefrontal cortex alters decision making during approach-avoidance conflict. *Soc Cogn Affect Neurosci* 2017;12:468-75.
31. Victor TA, Khalsa SS, Simmons WK, et al. Tulsa 1000: a naturalistic study protocol for multilevel assessment and outcome prediction in a large psychiatric sample. *BMJ Open* 2018;8:e016620.
32. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 2010;167:748-51.
33. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606-13.
34. Norman SB, Hami Cissell S, Means-Christensen AJ, et al. Development and validation of an overall anxiety severity and impairment scale (OASIS). *Depress Anxiety* 2006;23:245-9.
35. Bohn M, Babor T, Kranzler H. Validity of the Drug Abuse Screening Test (DAST-10) in inpatient substance abusers. *Problems Drug Depend* 1991;119:233-5.
36. Aupperle RL, Paulus MP, Kuplicki R, et al. Web-based graphic representation of the life course of mental health: cross-sectional study across the spectrum of mood, anxiety, eating, and substance use disorders. *JMIR Ment Health* 2020;7:e16919.
37. Misaki M, Tsuchiyagaito A, Al Zoubi O, et al. Connectome-wide search for functional connectivity locus associated with pathological rumination as a target for real-time fMRI neurofeedback intervention. *Neuroimage Clin* 2020;26:102244.
38. Ekhtiari H, Kuplicki R, Yeh HW, et al. Physical characteristics not psychological state or trait characteristics predict motion during resting state fMRI. *Sci Rep* 2019;9:419.
39. Stewart JL, Khalsa SS, Kuplicki R, et al. Interoceptive attention in opioid and stimulant use disorder. *Addict Biol* 2019;e12831.
40. Feng C, Forthman KL, Kuplicki R, et al. Neighborhood affluence is not associated with positive and negative valence processing in adults with mood and anxiety disorders: a Bayesian inference approach. *Neuroimage Clin* 2019;22:101738.
41. Le TT, Kuplicki RT, McKinney BA, et al. A nonlinear simulation framework supports adjusting for age when analyzing brainAGE. *Front Aging Neurosci* 2018;10:317.
42. Huang H, Thompson W, Paulus MP. Computational dysfunctions in anxiety: failure to differentiate signal from noise. *Biol Psychiatry* 2017;82:440-6.
43. Ford BN, Yolken RH, Aupperle RL, et al. Association of early-life stress with cytomegalovirus infection in adults with major depressive disorder. *JAMA Psychiatry* 2019;76:545-7.
44. Clausen AN, Aupperle RL, Yeh HW, et al. Machine learning analysis of the relationships between gray matter volume and childhood trauma in a transdiagnostic community-based sample. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2019;4:734-42.
45. Al Zoubi O, Mayeli A, Tsuchiyagaito A, et al. EEG microstates temporal dynamics differentiate individuals with mood and anxiety disorders from healthy subjects. *Front Hum Neurosci* 2019;13:56.
46. Al Zoubi O, Ki Wong C, Kuplicki RT, et al. Predicting age from brain EEG signals—a machine learning approach. *Front Aging Neurosci* 2018;10:184.
47. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59(Suppl 20):22-33, quiz 4-57.
48. Lang PJ, Bradley MM, Cuthbert BN. *International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical report A-8*. Gainsville (FL): University of Florida; 2008.
49. Lang P, Bradley M. *The international affective digitized sounds (2nd edition; IADS-2: affective ratings of sounds and instruction manual*. Gainsville (FL): University of Florida; 2007.
50. Peters J, Bromberg U, Schneider S, et al. Lower ventral striatal activation during reward anticipation in adolescent smokers. *Am J Psychiatry* 2011;168:540-9.
51. Peters J, Büchel C. Neural representations of subjective reward value. *Behav Brain Res* 2010;213:135-41.
52. Friston KJ, Parr T, de Vries B. The graphical brain: belief propagation and active inference. *Network Neuroscience*. 2017;1:381-414.
53. Parr T, Friston K. Working memory, attention, and salience in active inference. *Sci Rep* 2017;7:14678.
54. Moritz S, Woodward TS. Jumping to conclusions in delusional and non-delusional schizophrenic patients. *Br J Clin Psychol* 2005;44: 193-207.
55. Huq SF, Garety PA, Hemsley DR. Probabilistic judgements in deluded and non-deluded subjects. *Q J Exp Psychol A* 1988;40:801-12.
56. FitzGerald TH, Schwartenbeck P, Moutoussis M, et al. Active inference, evidence accumulation, and the urn task. *Neural Comput* 2015;27:306-28.
57. Friston KJ, Lin M, Frith C, et al. Active inference, curiosity and insight. *Neural Comput* 2017;29:2633-83.
58. Daunizeau J, den Ouden HE, Pessiglione M, et al. Observing the observer (I): meta-bayesian models of learning and decision-making. *PLoS One* 2010;5:e15554.
59. Gershman S. What does the free energy principle tell us about the brain? *arXiv* 2019:1901.07945.
60. Friston K, Mattout J, Trujillo-Barreto N, et al. Variational free energy and the Laplace approximation. *Neuroimage* 2007;34:220-34.
61. Rigoux L, Stephan KE, Friston KJ, et al. Bayesian model selection for group studies—revisited. *Neuroimage* 2014;84:971-85.
62. Stephan KE, Penny WD, Daunizeau J, et al. Bayesian model selection for group studies. *Neuroimage* 2009;46:1004-17.
63. Johnstone B, Callahan CD, Kapila CJ, et al. The comparability of the WRAT-R reading test and NAART as estimates of premorbid intelligence in neurologically impaired patients. *Arch Clin Neuropsychol* 1996;11:513-9.
64. Miller GA, Chapman JP. Misunderstanding analysis of covariance. *J Abnorm Psychol* 2001;110:40-8.
65. Mahoney JJ, Kalechstein AD, De Marco AP, et al. The relationship between premorbid IQ and neurocognitive functioning in individuals with cocaine use disorders. *Neuropsychology* 2017;31:311-8.
66. Altemus M, Sarvaiya N, Neill Epperson C. Sex differences in anxiety and depression clinical perspectives. *Front Neuroendocrinol*. 2014;35:320-30.
67. Koenen KC, Moffitt TE, Roberts AL, et al. Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *Am J Psychiatry* 2009;166:50-7.
68. Bjelland I, Krokstad S, Mykletun A, et al. Does a higher educational level protect against anxiety and depression? The HUNT study. *Soc Sci Med* 2008;66:1334-45.
69. Bekker MH, van Mens-Verhulst J. Anxiety disorders: sex differences in prevalence, degree, and background, but gender-neutral treatment. *Gend Med* 2007;4 Suppl B:S178-93.

70. Zammit S, Allebeck P, David AS, et al. A longitudinal study of premorbid IQ score and risk of developing schizophrenia, bipolar disorder, severe depression, and other nonaffective psychoses. *Arch Gen Psychiatry* 2004;61:354-60.
71. Kubicka L, Matejcek Z, Dytrych Z, et al. IQ and personality traits assessed in childhood as predictors of drinking and smoking behaviour in middle-aged adults: a 24-year follow-up study. *Addiction* 2001;96:1615-28.
72. Gater R, Tansella M, Korten A, et al. Sex differences in the prevalence and detection of depressive and anxiety disorders in general health care settings: report from the World Health Organization Collaborative Study on Psychological Problems in General Health Care. *Arch Gen Psychiatry* 1998;55:405-13.
73. DeVille DC, Kuplicki R, Stewart JL, et al. Diminished responses to bodily threat and blunted interoception in suicide attempters. *Elife* 2020;9:e51593.
74. Stewart JL, May AC, Poppa T, et al. You are the danger: attenuated insula response in methamphetamine users during aversive interoceptive decision-making. *Drug Alcohol Depend* 2014;142:110-9.
75. Hester R, Bell RP, Foxe JJ, et al. The influence of monetary punishment on cognitive control in abstinent cocaine-users. *Drug Alcohol Depend* 2013;133:86-93.
76. Simons JS, Dvorak RD, Batien BD. Methamphetamine use in a rural college population: associations with marijuana use, sensitivity to punishment, and sensitivity to reward. *Psychol Addict Behav* 2008;22:444-9.
77. Simons JS, Arens AM. Moderating effects of sensitivity to punishment and sensitivity to reward on associations between marijuana effect expectancies and use. *Psychol Addict Behav* 2007;21:409-14.
78. Smith R, Schwartenbeck P, Stewart JL, et al. Imprecise action selection in substance use disorder: evidence for active learning impairments when solving the explore-exploit dilemma. *Drug Alcohol Depend* 2020;215:108208.
79. Wiers RW, Eberl C, Rinck M, et al. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychol Sci* 2011;22:490-7.
80. Clark J, Watson S, Friston K. What is mood? A computational perspective. *Psychol Med* 2018;48:2277-84.
81. Peters A, McEwen BS, Friston K. Uncertainty and stress: why it causes diseases and how it is mastered by the brain. *Prog Neurobiol* 2017;156:164-88.
82. Joffily M, Coricelli G. Emotional valence and the free-energy principle. *PLOS Comput Biol* 2013;9:e1003094.
83. Smith R, Lane RD, Parr T, et al. Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance. *Neurosci Biobehav Rev* 2019;106:473-91.
84. Smith R, Parr T, Friston KJ. Simulating emotions: an active inference model of emotional state inference and emotion concept learning. *Front Psychol* 2019;10:2844.
85. de Berker AO, Rutledge R, Mathys C, et al. Computations of uncertainty mediate acute stress responses in humans. *Nat Commun* 2016;29:10996.
86. Hesp C, Smith R, Allen M, et al. Deeply felt affect: the emergence of valence in deep active inference. *Neural Comput* 2020. In press.
87. FitzGerald TH, Dolan RJ, Friston K. Dopamine, reward learning, and active inference. *Front Comput Neurosci* 2015;9:136.
88. Friston K, Schwartenbeck P, FitzGerald T, et al. The anatomy of choice: dopamine and decision-making. *Philos Trans R Soc Lond B Biol Sci* 2014;369:20130481.
89. Friston KJ, Shiner T, FitzGerald T, et al. Dopamine, affordance and active inference. *PLoS Comput Biol* 2012;8:e1002327.
90. Schwartenbeck P, FitzGerald TH, Mathys C, et al. The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cereb Cortex* 2015;25:3434-45.
91. Huys QJ, Tobler P, Hasler G, et al. The role of learning-related dopamine signals in addiction vulnerability. *Prog Brain Res* 2014; 211:31-77.
92. Koob GF, Volkow N. Neurocircuitry of addiction. *Neuropsychopharmacology* 2010;35:217-38.