A computational theory of evaluation processes in apathy

Francesco Rigoli[1*], Cristina Martinelli[2**]

[1] *City, University of London, Northampton Square, London, EC1V 0HB, UK*

[3]*Kingston University, Penrhyn Road, Kingston Upon Thames, Surrey, KT1 2EE, UK*

Correspondence: Francesco Rigoli

Department of Psychology

City, University of London

Northampton Square, London, UK EC1V 0HB

francesco.rigoli@city.ac.uk

Authors' contributions: F.R. developed the study concept. F.R. and C.R. collected and analysed the data and wrote the manuscript.

A computational theory of evaluation processes in apathy

Abstract

Computational modelling can offer valuable insight on mental illness. However, this approach has rarely been adopted to investigate apathy, a condition characterising a variety of psychiatric and neurological syndromes. This paper proposes a computational model of apathy and tests key model predictions in the healthy adult population. Building upon recent reference-dependent theories of evaluation, the model interprets apathy as arising from an excessive uncertainty about the distribution of incentives in the environment. This predicts that high-apathy individuals appraise the value of stimuli as less extreme and as more similar to one another. These predictions were assessed in two online studies where healthy adults rated the value of pictures characterised by varying levels of emotional salience. In line with the model, we observed that high-apathy individuals perceive negative stimuli as less negative, positive stimuli as less positive, and discriminate less among stimuli characterised by different salience. The contribution of this paper is twofold. On a more specific level, it sheds light on the precise mechanisms underlying evaluation processes in apathy. On a more general level, it highlights the insight offered by models of reference-dependent evaluation for understanding psychopathology.

Keywords: apathy, computational, reference dependent, uncertainty, evaluation

1. Introduction

Apathy describes a persistent state of lack of motivation and of blunted emotional reactivity (Husain & Roiser, 2018; Marin, 1991). This condition is common in psychiatric disorders such as depression (Yuen et al., 2015) and schizophrenia (Brown & Pluck, 2000; Yazbek, 2014), as well as in various neurological disorders including those produced by stroke (Caeiro et al., 2013) and traumatic brain injury (Starkstein & Pahissa, 2014), Alzheimer's disease (Zhao et al., 2016), Parkinson's disease (den Brok et al., 2015), vascular dementia (Staekenborg et al., 2010), Huntington's disease (van Duijn, 2014), and frontotemporal dementia (Chow et al., 2009). Moreover, evidence indicates that, to some degree, apathic tendencies exist also within the non-clinical population (Ang et al., 2017).

At the root of apathy are believed to be impairments in the fronto-striatal brain circuit (Husain & Roiser, 2018), due to which stimuli in the environment lose their salience and thus fail to elicit any emotional reaction or approach/avoidance drive. As a consequence, the typical signs of apathy are thought to arise, including behavioural inhibition, lack of initiative, and reduced emotional reactivity. This picture distinguishes apathy from other psychopathological conditions such as anxiety or low mood (Marin, 1991). For example, although both apathy and anxiety produce behavioural inhibition, the latter reflects distinct processes in the two conditions, that is, it reflects active avoidance of dangerous stimuli in anxiety and lack of engagement in apathy.

Recently, it has been argued that a computational approach can offer valuable insight on mental disorders (Huys et al., 2016; Montague et al., 2012; Valeri, 2021) such as apathy (Hezemans et al., 2020; Husain & Roiser, 2018). By adopting mathematical modelling, this approach can offer a precise description of the computational processes characterising a disorder. However, the computational processes underlying apathy remain largely to be explored (Hezemans et al., 2020; Husain & Roiser, 2018). This paper aims to fill this gap by proposing a computational model of apathy. After overviewing the model, the paper reports two empirical studies where key model predictions were tested empirically.

2. The model

Central to contemporary accounts of adaptive behaviour (Glimcher & Fehr, 2013), the notion of evaluation is at the core of our proposal. During evaluation, the brain assigns positive or negative value to surrounding stimuli, eliciting an approach or avoidance tendency, respectively. This perspective interprets "hot" aspects of cognition such as emotion, mood, affect, and decision-making as manifestations of evaluation processes. Given that, as discussed above, aberrations in such "hot" domains are at the core of apathy (Husain & Roiser, 2018), our proposal interprets this disorder as a form of impaired evaluation.

What does research know about evaluation processes? Recent perspectives highlight the reference-dependent nature of evaluation (Louie et al., 2013; Rigoli, 2019; Rigoli & Pezzulo, 2022; Stewart et al., 2016): the idea is that outcomes are not evaluated in isolation, but relative to their context. As an example, consider an individual who, while purchasing a house, discovers that the price is £10 more than expected. Compare this with someone paying for a coffee and realising that the price is £10 more than expected. Objectively, both individuals experience an unforeseen extra-cost of £10. Yet, the second person will arguably be way more upset. This example stresses the idea that evaluation is reference-dependent, namely, the notion that the subjective value of outcomes is not absolute, but relative. Recent work has started to explore the implications of this for understanding mental illness (Rigoli, 2022; Rigoli & Martinelli, 2021; Rigoli et al., 2021). Following this literature, we propose to interpret apathy as a form of aberrant reference-dependent evaluation; we refer to our proposal as to the Reference Dependent Model of Apathy (RDMA).

To introduce the RDMA (Rigoli, 2021; Rigoli, 2022; Rigoli et al., 2021), consider an environment or context (e.g., school) where a set of outcomes (e.g., school marks) can be experienced, each associated with a raw value (e.g., the actual mark). For each outcome, the calculation of the subjective value V_R associated with the raw value R depends on the following logistic function:

$$V_{R,k} = \frac{1}{1 + e^{-\frac{R-\mu}{\sigma}}} \quad (1)$$

The parameter μ reflects the reference point, and the parameter σ indicates the uncertainty about outcomes. The former corresponds to a standard to which outcomes are compered: R > μ implies $V_R > 0.5$ and a reward experience, R < μ implies $V_R < 0.5$ and punishment, and R = μ implies $V_R = 0.5$ and a neutral experience. For instance, the reference point μ might indicate the standard during a championship: an outcome better than the standard would be perceived as success, one worse than the standard as failure. The uncertainty parameter σ reflects the weight attributed to a discrepancy from the reference point. With high uncertainty, a discrepancy is weighted little; for example, an outcome above/below the standard is not considered too good/bad. Conversely, with low uncertainty, a discrepancy is weighted heavily; the same outcome above/below the standard is considered as very good/bad. Altogether, this captures the notion that subjective value is not absolute, but reference dependent.

Within this framework, adaptive evaluation occurs when an individual has a correct representation of the distribution of outcomes in the environment, hence possessing appropriate parameters (Rigoli et al., 2021). For example, consider an environment where the average raw value is 0 and the SD is 20 (fig. 1). Adaptive evaluation ensues when outcomes are assessed based on μ = 0 and σ = 20, reflecting the true statistics. Conversely, evaluation is maladaptive when any of these parameters is altered. Specific parameter alterations can be linked with specific clinical manifestations; for instance, an excessive standard parameter μ has been proposed to underpin anorexia nervosa and depression (Rigoli, 2022; Rigoli & Martinelli, 2021; Rigoli et al., 2021).

The RDMA argues that apathy emerges when the uncertainty parameter σ is excessive, that is, much larger than the actual environmental SD (fig. 1). In our example, apathy can be described by an uncertainty parameter σ = 40 (remember that, in this example, the SD is equal to 20). What is the consequence of such inflated uncertainty parameter? To answer this, consider four possible outcomes: -40, -20, 20, and 40 (fig. 1). Comparing adaptive evaluation (where σ = 20) against apathy (where σ = 40), in both cases the subjective value V_R for -40 and -20 is smaller than 0.5, indicating punishment; while for 40 and 20 it is larger than 0.5, indicating reward. However, comparing again the two conditions, the subjective value V_R for -40 and -20 is lower, whereas it is higher for 40 and 20. In other

words, during adaptive evaluation compared to apathy, subjective values appear as more distant from neutrality (occurring when V_R=0.5). Moreover, comparing the two conditions, the distance in subjective value among outcomes is magnified: this is true for the distance between -40 and -20, the distance between -20 and 20, and between 20 and 40.

Overall, the RDMA interprets apathy as a form of abnormal reference dependent evaluation (Louie et al., 2013; Rigoli, 2019; Stewart et al., 2016); specifically, as ensuing from an exaggerated uncertainty parameter σ. Two main predictions arise from this proposal. First, apathetic individuals are predicted to assess outcomes as more neutral (i.e., less extreme in value). Second, these individuals are predicted to discriminate less among stimuli, that is, to view stimuli as more similar to one another in terms of value. Do these predictions fit with available knowledge about apathy? The prediction that outcomes become more neutral is broadly consistent with the definition of the disorder. Consider key features such as neutral affective and motivational stance, poor motivation to approach or avoid outcomes, and inhibited emotional reactions. All these features are consistent with the RDMA prediction that outcomes become more neutral. The second model prediction is consistent with recent empirical evidence showing that, during decision-making, apathy measures correlate with the ability to distinguish between monetary reward and no reward (Le Bouc et al., 2016). An impaired ability to make this distinction is consistent with the RDMA prediction that apathy is characterised by diminished discrimination ability.

Although, as we have just seen, the RDMA is broadly consistent with available evidence, the model remains to be assessed in a systematic fashion. Here we do this in two empirical investigations. To explain the rationale adopted, consider the example above but now where the raw values correspond to general labels such as Highly Negative, Mildly negative, Mildly positive, and Highly Positive, instead of -40, -20, 20, and 40, respectively (fig. 2A). Here the RDMA makes three predictions. First, high-apathy individuals are predicted to evaluate both Mildly positive and Mildly negative stimuli as more neutral and as more similar to one another (fig. 2B). Second, high-apathy individuals are predicted to evaluate Highly Negative and Mildly negative stimuli as less negative and as more similar to one another (fig. 2C). Third, high-apathy individuals are predicted to evaluate Highly positive and Mildly

positive stimuli as less positive and more similar to one another. Prediction one was tested in Study 1; Study 2 aimed at replicating Study 1 and at testing prediction two and three.

3. Study 1

Healthy adult participants were presented with pictures taken from Oasis (Kurdi et al., 2017), a database standardised in terms of perceived positive or negative value. Based on the database, half of the pictures presented to participants were associated with Mildly positive value, and the other half with Mildly negative value. Participants were asked to rate pictures on a 7-point scale ranging from Very negative to Very positive. After pictures were presented, participants filled the Apathy Motivation Index (AMI) (Ang et al., 2017), a questionnaire of self-reported apathy validated in the healthy population. Based on their reported apathy score, participants were labelled as high-apathy or low-apathy (grouping was based on a median split). This allowed us to analyse participants' ratings based on a 2x2 mixed ANOVA with Value (Mildly positive vs Mildly negative; these were based on calculating, for each participant, the average score for each Value category) as within-subjects factor and Apathy (low-apathy vs high-apathy) as between-subjects factor. The RDMA predicts that this analysis revealed (i) an interaction between Apathy and Value, indicating high-apathy individuals reporting lower score for Mildly positive pictures and higher score for Mildly negative pictures, and (ii) no main effect of Apathy (fig. 2B, 3A) (not of interest here, a main effect of Value was also expected).

3.1 Participants

Participants were recruited online via the Prolific website (ww.prolific.co). The sample included 194 adults (mean age 35, 105 females; all participants recruited were included in the analysis) from the UK reporting absence of any mental health issue or any medication affecting the central nervous system. The sample size was established a priori based on a power analysis performed on G-power (Faul et al., 2007) (assuming Type-1 error probability $\alpha = .05$, power $\beta = .9$, and effect size $\eta^2 = .2$). The study was approved by the Ethics Committee of the university where the study was conducted.

3.2 Stimuli and measures

Participants were presented with pictures from the Oasis database (Kurdi et al., 2017). This includes images of various kind (e.g., places, objects, people etc.), each associated with a score ranging from 1 (Very negative) to 2 (Moderately Negative), 3 (Somewhat negative), 4 (Neutral), 5 (Somewhat positive), 6 (Moderately positive), and 7 (Very positive). Each score reflects the rate about the value of the picture averaged across people who participated in the original study where the database was validated (Kurdi et al., 2017). Here, from the Oasis pool, we selected 20 pictures associated with a Mildly negative score (ranging from 2.7 to 3.2) and 20 pictures associated with a Mildly positive score (ranging from 4.8 to 5.3).

To measure apathy, the AMI questionnaire was administered (Ang et al., 2017). For each of 18 items, this asks participants to indicate how much they agree (on a scale ranging from "completely untrue" to "mostly untrue", "neither true nor untrue", "mostly true", and "completely true) with a statement (e.g., "I do not like to laze around"; note that, like in this example, all statements are framed in a way that the highest apathy level is expressed by indicating "completely untrue"). The AMI isolates three distinct dimensions including behavioural ($Apathy_{BEH}$), social ($Apathy_{SOC}$), and emotional ($Apathy_{EMO}$) apathy (each associated with 6 items). $Apathy_{BEH}$ indicates lack of motivation, initiative, and activity, as captured by items such as "I do not like to laze around". $Apathy_{SOC}$ reflects a passive disposition towards social interactions, as captured by items such as "I start conversation with random people". $Apathy_{EMO}$ describes inhibited emotional reactions to salient event, as captured by items such as "I feel sad or upset when I hear bad news". Weak or no correlation among AMI dimensions have been reported. Regarding reliability, analysis of internal consistency has revealed a Cronbach's alpha of .79 for $Apathy_{BEH}$, .75 for $Apathy_{SOC}$, and .75 for $Apathy_{EMO}$, with test-retest reliability being .88 for $Apathy_{BEH}$, .84 for $Apathy_{SOC}$, and .72 for $Apathy_{EMO}$ (Ang et al., 2017). Supporting the validity of AMI, each dimension exhibits substantial correlations with appropriate scales of other questionnaires measuring apathy and related constructs (Ang et al., 2017).

Because we did not have any a priori hypothesis about which AMI dimension might impact upon participants' ratings, we run three separate ANOVA analysis, each including one AMI dimension as between-subject factor (with the within-subject factor being Value in all ANOVAs). Note that, for the ANOVAs, each AMI dimension was dichotomised based on a median split, thus separating low-apathy and high-apathy participants.

3.3 Procedure

After participants accepted to take part, Prolific directed them to an online study run via the Gorilla software (www.gorilla.sc). Following task instruction, a sequence of 40 pictures was displayed in random order. Every time a new picture appeared, participants indicated its value choosing among the following options: 1 (Very negative), 2 (Moderately Negative), 3 (Somewhat negative), 4 (Neutral), 5 (Somewhat positive), 6 (Moderately positive), and 7 (Very positive). The rating was finalised by a mouse click (with no time limit), after which a new picture was immediately presented. After completing the task, participants filled an online version of the AMI questionnaire. Completing the study took approximately 5 minutes and was rewarded with a £.50.

3.4 Results

Across participants, the median score (based on which participants were grouped) for $Apathy_{BEH}$, $Apathy_{SOC}$, and $Apathy_{EMO}$ was 5, 10, and 12, respectively. The RDMA predicts that the ANOVA of participants' ratings will reveal (i) a Value-Apathy interaction (showing high- compared to low-apathy participants reporting higher ratings for Mildly negative images and lower ratings for Mildly positive images) and (ii) no main effect of Apathy (fig. 2B, 3A). These predictions fit with observations when $Apathy_{BEH}$ was considered as between-subject factor (tab. 1; fig. 3B): besides a main effect of Value ($F(1,192) = 1958.96$, $p < .001$, $\eta_P^2 = .91$), a Value-$Apathy_{BEH}$ interaction ($F(1,192) = 10.15$, $p = .002$, $\eta_P^2 = .05$) emerged together with no main effect of $Apathy_{BEH}$ (($F(1,192) = .61$, $p = .437$, $\eta_P^2 < .01$). The interaction effect indicated that, compared to low-$Apathy_{BEH}$ individuals, high-$Apathy_{BEH}$

participants exhibited higher ratings for Mildly negative images and lower ratings for Mildly positive images.

However, results did not fit with RDMA predictions when $Apathy_{SOC}$ was considered as between-subjects factor (tab. 1; fig. 3C): although no main effect of $Apathy_{SOC}$ emerged ((F(1,192) = .88, p = .350, $\eta_P^2 < .01$), no Value-$Apathy_{SOC}$ interaction emerged either ((F(1,192) = 1.04, p = .308, $\eta_P^2 < .01$). Likewise, results failed to support RDMA predictions when $Apathy_{EMO}$ was considered as between-subject factor (tab. 1; fig. 3D) because, although no main effect of $Apathy_{EMO}$ emerged ((F(1,192) = 1.38, p = .242, $\eta_P^2 < .01$), no Value-$Apathy_{EMO}$ interaction emerged either ((F(1,192) = .37, p = .544, $\eta_P^2 < .01$)

Altogether, Study 1 supports RDMA predictions when apathy is defined in terms of $Apathy_{BEH}$, but not in terms of $Apathy_{SOC}$ nor $Apathy_{EMO}$; thus, $Apathy_{SOC}$ and $Apathy_{EMO}$ were not examined further. Focusing exclusively on $Apathy_{BEH}$, Study 2 assessed other central RDMA predictions.

4. Study 2

For Study 2, participants performed three separate tasks. The first task was exactly the same as in Study 1, with the aim to replicate it. The second task adopted a similar structure, but now employing pictures from the Oasis database (Kurdi et al., 2017) associated with either Highly negative or Mildly negative score. The third task was also similar, but now employing pictures from the Oasis database associated with either Highly positive or Mildly positive score. Each task was analysed with a 2x2 mixed ANOVA of participants' ratings. All ANOVAs had $Apathy_{BEH}$ (established by a median split) as between-subject factor; the within-subjects Value factor opposed Mildly positive versus Mildly negative images for task one, Highly negative versus Mildly negative images for task two, and Highly positive versus Mildly positive images for task three.

Regarding RDMA predictions, for task one the theory implies the same predictions as in Study 1: (i) an $Apathy_{BEH}$-Value interaction, indicating low-$Apathy_{BEH}$ individuals reporting lower score for Mildly

positive pictures and higher score for Mildly negative pictures, and (ii) no main effect of $Apathy_{BEH}$ (fig. 4A). For task two, the RDMA predicts (i) a main effect of $Apathy_{BEH}$, reflecting overall higher ratings expressed by high-$Apathy_{BEH}$ individuals; and (ii) an $Apathy_{BEH}$-Value interaction indicating that high-$Apathy_{BEH}$ individuals rate Highly negative and Mildly negative pictures as more similar (fig. 2C; 4C). For task three, the RDMA predicts (i) a main effect of $Apathy_{BEH}$, reflecting overall lower ratings expressed by high-$Apathy_{BEH}$ individuals; and (ii) an $Apathy_{BEH}$-Value interaction indicating that high-$Apathy_{BEH}$ individuals rate Highly positive and Mildly positive pictures as more similar (fig. 2D; 4E).

4.1 Participants

Participants were recruited online via the Prolific website (ww.prolific.co). The sample included a new pool of 194 adults (mean age 34, 113 females; all participants recruited were included in analyses) from the UK reporting absence of any mental health issue or any medication affecting the central nervous system. The sample size was established a priori as in Study 1. The study was approved by the Ethics Committee of the university where the study was conducted.

4.2 Stimuli and measures

For task one, the same stimuli from the Oasis database used in Study 1 were adopted. For study two, different stimuli were selected, 20 associated with a Mildly negative score (ranging from 2.7 to 3.2) and 20 with a Highly negative score (ranging from 1.9 to 2.2). For study three, 20 stimuli were associated with a Mildly positive score (ranging from 4.8 to 5.3) and 20 with a Highly positive score (ranging from 5.8 to 6.1). As in Study 1, the AMI was administered, now with a specific focus on $Apathy_{BEH}$. This variable was dichotomised based on a median split, thus separating low-$Apathy_{BEH}$ and high-$Apathy_{BEH}$ participants.

4.3 Procedure

After participants accepted to take part, Prolific directed them to an online study run by the Gorilla software (www.gorilla.sc). Following instructions, participants performed task one followed by task two and three, always in this order. For task one, every time a new picture appeared, participants indicated its value choosing among the following options: 1 (Very negative), 2 (Moderately Negative), 3 (Somewhat negative), 4 (Neutral), 5 (Somewhat positive), 6 (Moderately positive), and 7 (Very positive). For task two, where only negative stimuli were presented, options were different: 1 (Extremely negative), 2 (Highly negative), 3 (Negative), 4 (Moderately negative), 5 (Somewhat negative), 6 (Neutral). For task three, where only positive stimuli were presented, options were also different: 1 (Neutral), 2 (Somewhat positive), 3 (Moderately positive), 4 (Positive), 5 (Highly positive), 6 (Extremely positive). In all tasks, the rating was finalised by a mouse click (with no time limit), after which a new picture was immediately presented. After completing the tasks, participants filled an online version of the AMI questionnaire. Completing the study took approximately 10 minutes and was rewarded with £1.

4.4 Results

Across participants, the median score (based on which participants were grouped) for $Apathy_{BEH}$ was 5. Replicating Study 1, the ANOVA for the first task (tab. 2; fig. 4B) revealed (besides a main effect of Value ($F(1,192) = 1787.35$, $p < .001$, $\eta_P^2 = .90$) (i) no main effect of $Apathy_{BEH}$ ($F(1,192) = 1.56$, $p = .214$, $\eta_P^2 < .01$) and (ii) a Value-$Apathy_{BEH}$ interaction ($F(1,192) = 6.36$, $p = .012$, $\eta_P^2 = .03$) indicating that, compared to low-$Apathy_{BEH}$ individuals, high-$Apathy_{BEH}$ participants exhibited higher ratings for Mildly negative images and lower ratings for Mildly positive images.

Regarding the ANOVA for task two (tab. 2; fig. 4D), results confirmed RDMA predictions by showing (besides a main effect of Value ($F(1,192) = 776.91$, $p < .001$, $\eta_P^2 = .80$)) (i) a main effect of $Apathy_{BEH}$ ($F(1,192) = 3.96$, $p = .048$, $\eta_P^2 = .02$), indicating that high-$Apathy_{BEH}$ participants reported higher

ratings, and (ii) a Value-$Apathy_{BEH}$ interaction (F(1,192) = 4.75, p = .030, η_P^2 = .02) indicating that high-$Apathy_{BEH}$ participants rated Highly negative and Mildly negative pictures as more similar.

Results were in line with RDMA predictions also for task three (tab. 2; fig. 4F), where the ANOVA revealed (besides a main effect of Value (F(1,192) = 836.34, p < .001, η_P^2 = .81)) (i) a main effect of $Apathy_{BEH}$ (though emerging only as a trend towards significance; F(1,192) = 3.69, p = .056, η_P^2 = .02), indicating that high-$Apathy_{BEH}$ participants reported lower ratings, and (ii) a Value-$Apathy_{BEH}$ interaction (F(1,192) = 5.67, p = .019, η_P^2 = .03), indicating that high-$Apathy_{BEH}$ participants rated Highly positive and Mildly positive pictures as more similar.

Altogether, while results for task one replicate Study 1, results for task two and three confirm other key RDMA predictions. In line with the theory, the overall picture emerging form these observations reveals that, in both positive and negative domains, apathy is characterised by attributing less extreme values to stimuli and by evaluating stimuli as more similar to one another.

5. Discussion

We propose the RDMA as a computational account of apathy, a condition frequent in psychiatric and neurological disorders (Husain & Roiser, 2018; Marin, 1991). Building on research about reference-dependent evaluation (Louie et al., 2013; Rigoli, 2019; Stewart et al., 2016), the theory interprets apathy as arising from an excessive uncertainty parameter σ. Two key predictions ensue: first, stimuli are predicted to appear as more neutral in value; second, the ability to discriminate stimuli in terms of value is predicted to be reduced. When testing these predictions empirically, we found supporting evidence: high-apathy individuals perceive negative images as less negative, positive images as less positive, and discriminate less among images with different value. Altogether, the picture offered by the RDMA sheds light on the computational mechanisms underlying key features of the disorder including behavioural inhibition, lack of initiative, and reduced emotional reactivity.

Research based on self-report questionnaires has pinpointed to partially independent apathy subtypes, including $Apathy_{BEH}$, $Apathy_{SOC}$, and $Apathy_{EMO}$ (Ang et al., 2017; Radakovic & Abraham, 2014). Our observations support the RDMA only regarding the first of these: $Apathy_{BEH}$. Lack of support regarding $Apathy_{SOC}$ does not appear as puzzling. Assume that $Apathy_{SOC}$ reflects a subcategory of $Apathy_{BEH}$, with $Apathy_{BEH}$ capturing lack of motivation broadly defined, and with $Apathy_{SOC}$ capturing lack of motivation specifically in social domains (this interpretation is supported by the presence of a correlation between the two dimensions; Ang et al., 2017). Being the RDMA a general account of apathy, and not specific for the social domain, it is not surprising that behavioural effects emerge for $Apathy_{BEH}$ but not for $Apathy_{SOC}$. Lack of support for the RDMA regarding $Apathy_{EMO}$ appears as more problematic at first: the concept of $Apathy_{EMO}$ is purportedly more pertinent for the RDMA. However, a careful scrutiny of how the AMI assesses $Apathy_{EMO}$ reveals that all items of this scale refer to negative emotions (with items such as "I feel sad or upset when I hear bad news"), and neglect positive emotions (Ang et al., 2017). This is problematic because the RDMA assumes that negative and positive emotions alike are affected by apathy – the AMI might thus be inappropriate to assess the RDMA with regard to $Apathy_{EMO}$. To our knowledge, self-report measures of $Apathy_{EMO}$ encompassing both negative and positive emotions are not available in the literature; developing such measures might reopen the possibility to assess the RDMA regarding $Apathy_{EMO}$.

Neuroimaging studies conducted among apathetic patients have reported abnormalities in specific brain regions encompassing the ventral striatum of the basal ganglia, the dopaminergic midbrain, the ventromedial prefrontal cortex, and the dorsal anterior cingulate cortex (Husain & Roiser, 2018). Notably, a large body of research indicates that this circuit underlies evaluation processes (Glimcher & Fehr, 2013; Kable & Glimcher, 2009; Rangel et al., 2008), consistent with the idea that apathy should be framed in terms of aberrant evaluation (Husain & Roiser, 2018). However, involvement of this circuit is not specific to apathy, but emerges also in other psychiatric conditions (Insel & Quirion, 2005). Thus, an important question is which specific impairments in these regions are associated with apathy and not with other conditions. The RDMA offers a possible answer. Research has shown that neural activity in these brain regions reflects the value of outcomes (or, similarly, a prediction error signal) (Glimcher &

Fehr, 2013; Kable & Glimcher, 2009; Rangel et al., 2008). On this basis, the RDMA predicts that apathy is characterised by (i) attenuated response in these regions for both rewards and punishments, and (ii) by a more similar response for rewards (or punishments) of different magnitude. Put another way, the RDMA predicts that, in apathy, the neural gain of brain regions implicated in evaluation is reduced. Partial support for this comes from evidence showing a decreased brain response to reward in schizophrenia, a syndrome characterised by apathy (Ziauddeen & Murray, 2010). A similar reasoning applies when considering the role of dopamine, a key neuromodulator implicated in evaluation (Wise, 2004). Research indicates that dopaminergic bursts from the dopaminergic midbrain to the ventral striatum reflect a reward prediction error signal (Glimcher & Fehr, 2013; Kable & Glimcher, 2009; Rangel et al., 2008; Schultz et al., 1997) (whether this signal integrates both reward and punishment information remains contentious; e.g., Rigoli et al., 2016). In this context, the RDMA predicts an attenuated dopaminergic signal in apathy, manifested as decreased response to reward and as a more similar response to rewards of different magnitude.

The RDMA raises the question of where an excessive uncertainty parameter, proposed as being at the core of apathy, comes from. Reference-dependent models of evaluation assume that the brain learns the parameters from experience (Louie et al., 2013; Rigoli, 2019; Stewart et al., 2016). On this basis, an excessive uncertainty parameter might result from experiencing a rapid alternation of extreme rewards and punishments. Genetic factors might also contribute, with some individuals being predisposed to develop an excessive uncertainty parameter. Finally, as the prevalence of apathy in traumatic brain injury and neurodegenerative disorders suggests, specific forms of brain damage might produce an excessive uncertainty parameter (Husain & Roiser, 2018). Exploring the mechanisms responsible of the development of an excessive uncertainty parameter represents a promising research avenue.

Finally, we highlight some limitations of our empirical investigation. First, our experiments employ visual images; whether our observations extend to other emotionally charged stimuli remains an open question. Second, we have adopted self-reports to measure evaluation; it remains to be investigated whether similar effects emerge also when implicit manifestations of evaluation, such as physiological responses, are considered. Third, our sample included healthy participants only. Although some healthy

individuals manifest high degrees of apathy, it remains debatable whether the same type of apathy characterises these individuals and people with a diagnosis (Ang et al, 2017); it is even far from certain whether different illnesses exhibit the same form of apathy, or whether different types of apathy can be recognised in different disorders (Marin, 1991). Testing the RDMA in clinical populations is paramount to assess the generality of the theory.

In summary, we propose the RDMA as a computational theory of apathy, aiming at translating clinical descriptions of this condition in the language of formal mathematical modelling. Besides clarifying key concepts in the literature, the RDMA makes specific predictions, some of which are tested here. In line with predictions, we observed that apathetic individuals evaluate positive pictures as less positive, negative pictures as less negative, and pictures as overall more similar to one another in terms of value. The contribution of this paper is twofold: on a more specific level, it sheds light on apathy, while on a more general level it highlights the insight offered by models of reference-dependent evaluation for understanding psychopathology.

Data availability statement

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Conflict of interest statement

The authors declare no conflict of interest.

References

Ang, Y. S., Lockwood, P., Apps, M. A., Muhammed, K., & Husain, M. (2017). Distinct subtypes of apathy revealed by the apathy motivation index. *PloS one*, *12*(1), e0169938.

Brown, R. G., & Pluck, G. (2000). Negative symptoms: the 'pathology'of motivation and goal-directed behaviour. *Trends in neurosciences*, *23*(9), 412-417.

Caeiro, L., Ferro, J. M., & Costa, J. (2013). Apathy secondary to stroke: a systematic review and meta-analysis. *Cerebrovascular Diseases*, *35*(1), 23-39.

Chow, T. W., Binns, M. A., Cummings, J. L., Lam, I., Black, S. E., Miller, B. L., ... & van Reekum, R. (2009). Apathy symptom profile and behavioral associations in frontotemporal dementia vs dementia of Alzheimer type. *Archives of Neurology*, *66*(7), 888-893.

den Brok, M. G., van Dalen, J. W., van Gool, W. A., Moll van Charante, E. P., de Bie, R. M., & Richard, E. (2015). Apathy in Parkinson's disease: a systematic review and meta-analysis. *Movement Disorders*, *30*(6), 759-769.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Glimcher, P. W., & Fehr, E. (Eds.). (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.

Hezemans, F. H., Wolpe, N., & Rowe, J. B. (2020). Apathy is associated with reduced precision of prior beliefs about action outcomes. *Journal of Experimental Psychology: General*.

Huys, Q. J., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*, *19*(3), 404.

Husain, M., & Roiser, J. P. (2018). Neuroscience of apathy and anhedonia: a transdiagnostic approach. *Nature Reviews Neuroscience*, *19*(8), 470-484.

Insel, T. R., & Quirion, R. (2005). Psychiatry as a clinical neuroscience discipline. *Jama*, *294*(17), 2221-2224.

Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, *63*(6), 733-745.

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the open affective standardized image set (OASIS). *Behavior research methods*, *49*(2), 457-470.

Le Bouc, R., Rigoux, L., Schmidt, L., Degos, B., Welter, M. L., Vidailhet, M., ... & Pessiglione, M. (2016). Computational dissection of dopamine motor and motivational functions in humans. *Journal of Neuroscience*, *36*(25), 6623-6633.

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, *110*(15), 6139-6144.

Marin, R. S. (1991). Apathy: a neuropsychiatric syndrome. *The Journal of neuropsychiatry and clinical neurosciences*.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in cognitive sciences*, *16*(1), 72-80.

Radakovic, R., & Abrahams, S. (2014). Developing a new apathy measurement scale: Dimensional Apathy Scale. *Psychiatry research*, *219*(3), 658-663.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience*, *9*(7), 545-556.

Rigoli, F. (2019). Reference effects on decision-making elicited by previous rewards. *Cognition*, *192*, 104034.

Rigoli, F. (2021). Political motivation: A referent evaluation mathematical model. *Journal of Social and Political Psychology*, *9*(1), 8-23.

Rigoli, F. (2022). When all glasses look half empty: a computational model of reference dependent evaluation to explain depression. *Journal of Cognitive Psychology*.

Rigoli, F., & Martinelli, C. (2021). A reference-dependent computational model of anorexia nervosa. *Cognitive, Affective, & Behavioral Neuroscience*, 1-9.

Rigoli, F., & Pezzulo, G. (2022). A reference-based theory of motivation and effort allocation. *Psychonomic Bulletin & Review*, 1-13.

Rigoli, F., Chew, B., Dayan, P., & Dolan, R. J. (2016). Multiple value signals in dopaminergic midbrain and their role in avoidance contexts. *Neuroimage*, *135*, 197-203.

Rigoli, F., Martinelli, C., & Pezzulo, G. (2021). The half-empty/full glass in mental health: a reference-dependent computational model of evaluation in psychopathology. *Clinical Psychological Science*, 2167702621998344.

Staekenborg, S. S., Su, T., van Straaten, E. C., Lane, R., Scheltens, P., Barkhof, F., & van der Flier, W. M. (2010). Behavioural and psychological symptoms in vascular dementia; differences between small-and large-vessel disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *81*(5), 547-551.

Starkstein, S. E., & Pahissa, J. (2014). Apathy following traumatic brain injury. *Psychiatric Clinics of North America*.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, *53*(1), 1-26.

Valeri, M. (2021). *Organizational Studies. Implications for the Strategic Management*. Springer, Switzerland

van Duijn, E., Craufurd, D., Hubers, A. A., Giltay, E. J., Bonelli, R., Rickards, H., ... & European Huntington's Disease Network Behavioural Phenotype Working Group. (2014). Neuropsychiatric symptoms in a European Huntington's disease cohort (REGISTRY). *Journal of Neurology, Neurosurgery & Psychiatry*, *85*(12), 1411-1418.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature reviews neuroscience*, *5*(6), 483-494.

Yazbek, H., Norton, J., Capdevielle, D., Larue, A., Boulenger, J. P., Gély-Nargeot, M. C., & Raffard, S. (2014). The Lille Apathy Rating Scale (LARS): exploring its psychometric properties in schizophrenia. *Schizophrenia research*, *157*(1-3), 278-284.

Yuen, G. S., Bhutani, S., Lucas, B. J., Gunning, F. M., AbdelMalak, B., Seirup, J. K., ... & Alexopoulos, G. S. (2015). Apathy in late-life depression: common, persistent, and disabling. *The American Journal of Geriatric Psychiatry*, *23*(5), 488-494.

Zhao, Q. F., Tan, L., Wang, H. F., Jiang, T., Tan, M. S., Tan, L., ... & Yu, J. T. (2016). The prevalence of neuropsychiatric symptoms in Alzheimer's disease: systematic review and meta-analysis. *Journal of affective disorders*, *190*, 264-271.

Ziauddeen, H., & Murray, G. K. (2010). The relevance of reward pathways for schizophrenia. *Current opinion in psychiatry*, *23*(2), 91-96.

Tab 1. Descriptive statistics of participants' ratings for different conditions in Study 1.

Group	Positive picture		Negative picture	
	Mean	**SD**	**Mean**	**SD**
Whole sample (n = 194)	5.04	.55	2.52	.61
high-$Apathy_{BEH}$ (n = 88)	4.97	.56	2.64	.58
low-$Apathy_{BEH}$ (n = 106)	5.10	.53	2.42	.61
high-$Apathy_{SOC}$ (n = 100)	5.02	.56	2.47	.56
low-$Apathy_{SOC}$ (n = 94)	5.06	.53	2.58	.66
high-$Apathy_{EMO}$ (n = 86)	4.98	.52	2.52	.61
low-$Apathy_{EMO}$ (n = 108)	5.09	.57	2.53	.61

Tab 2. Descriptive statistics of participants' ratings for different tasks and conditions in Study 2.

Group	Task one				Task two				Task three			
	Positive picture		Negative picture		Highly negative		Mildly negative		Highly positive		Mildly positive	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Whole sample (n = 194)	5.09	.62	2.47	.60	2.88	.89	4.15	.99	4.14	.99	2.61	.91
high-$Apathy_{BEH}$ (n = 90)	5.05	.70	2.58	.65	3.04	.96	4.22	1.09	3.97	1.05	2.56	.97
low-$Apathy_{BEH}$ (n = 104)	5.13	.52	2.34	.49	2.69	.77	4.07	.86	4.33	.88	2.68	.84

Fig. 1. Description of the RDMA

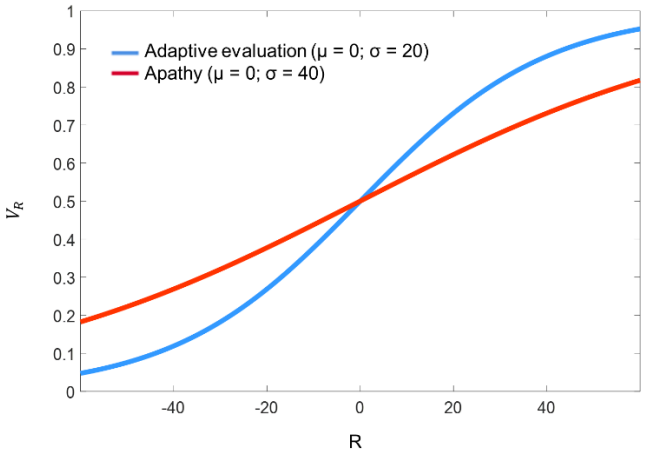

Fig. 2. Predictions of the RDMA predictions and results considering different AMI dimensions

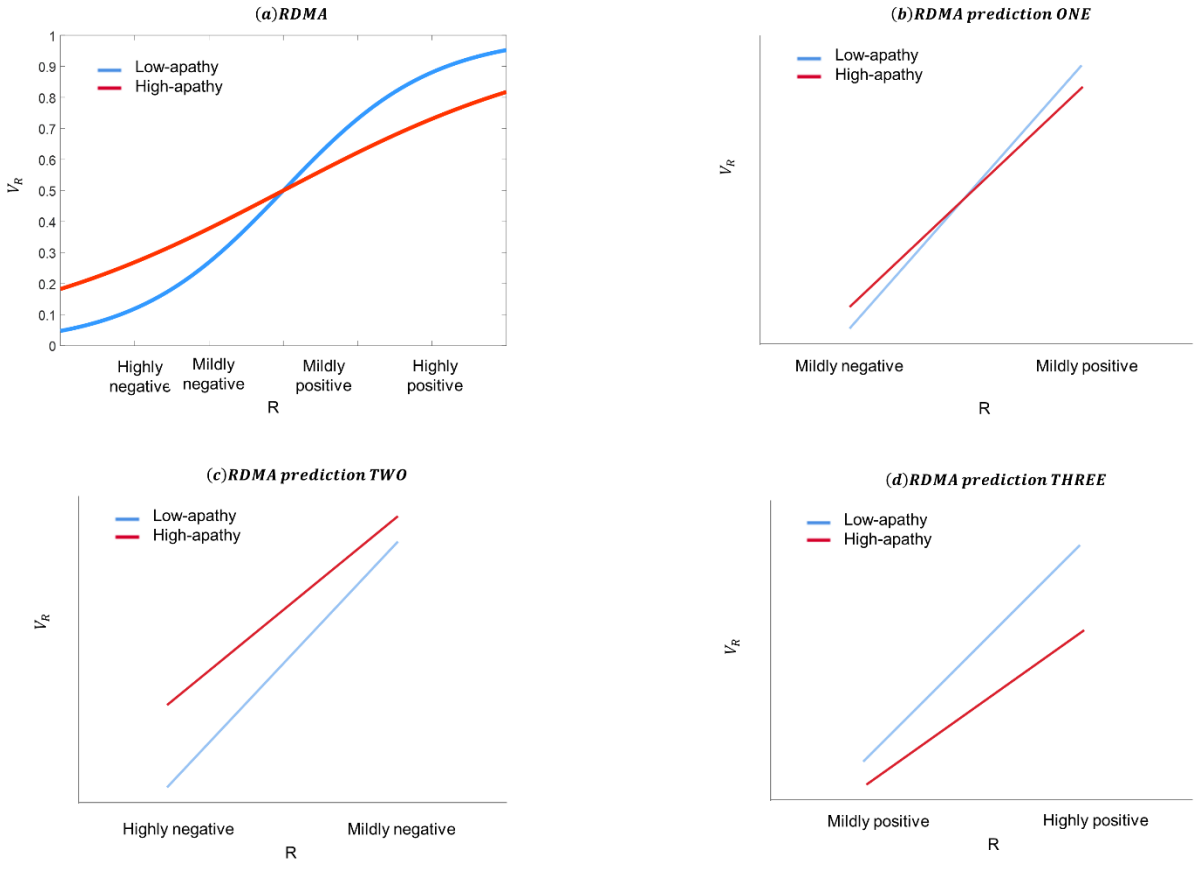

Fig. 3. Study 1: RDMA predictions and results considering different AMI dimensions

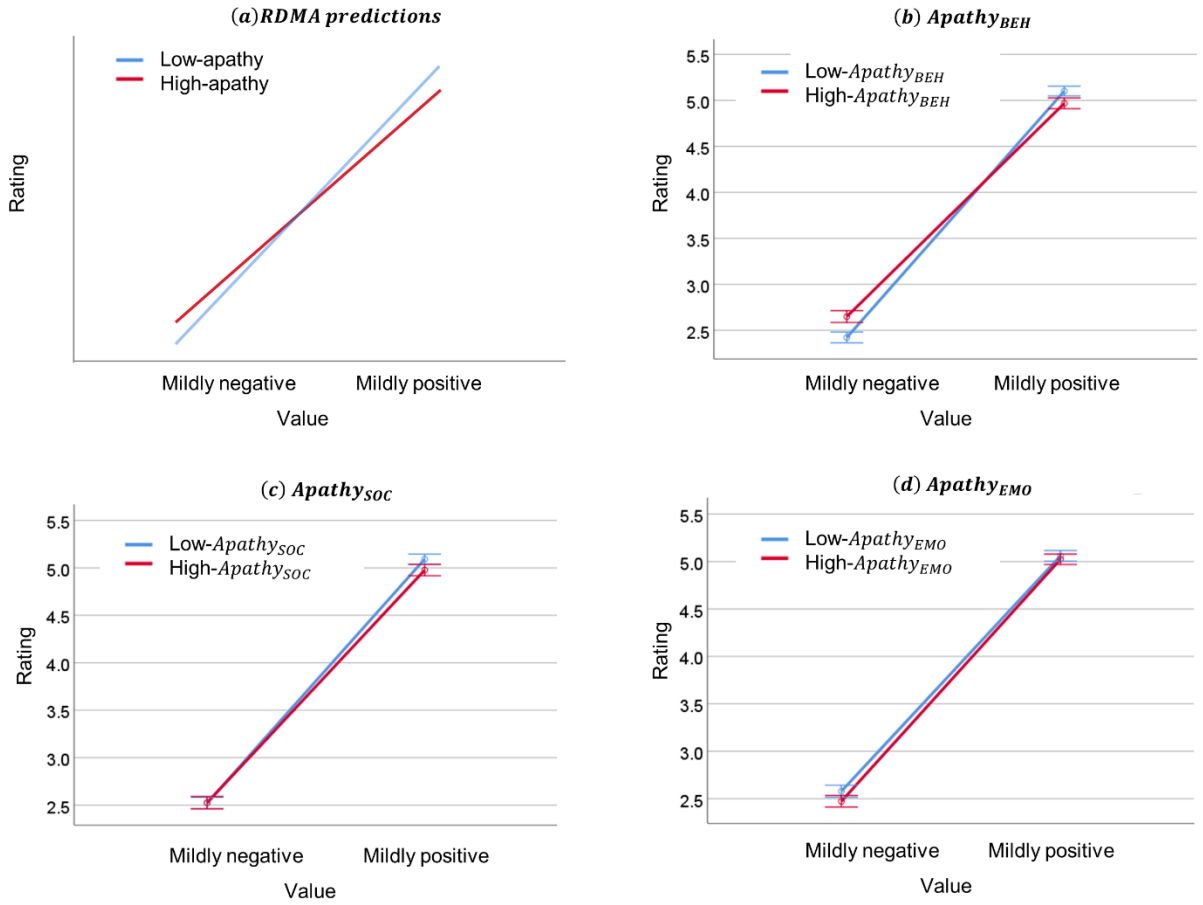

Fig. 4. Study 2: RDMA predictions and results for different tasks