*Theoretical/Methodological/Review Article*

# The Half-Empty/Full Glass in Mental Health: A Reference-Dependent Computational Model of Evaluation in Psychopathology

## Francesco Rigoli[1] , Cristina Martinelli[2,3], and Giovanni Pezzulo[4]

[1]Department of Psychology, City, University of London; [2]Institute of Psychiatry, Psychology & Neuroscience, King's College London; [3]Department of Psychology, Kingston University; and [4]Insitute of Cognitive Sciences and Technologies, National Research Council of Italy, Rome, Italy

## Abstract

Evaluation (the process attributing value to outcomes) underlies "hot" aspects of cognition, such as emotion, affect, and motivation. In several psychopathologies, such as depression and addiction, impairments in evaluation are critical. Contemporary theories highlight the reference-dependent nature of evaluation, whereby outcomes are evaluated relative to their context. Surprisingly, reference-dependent evaluation remains to be explored in the context of psychopathology. We offer a computational theory of how impaired reference-dependent evaluation might underlie mental illness. The theory proposes that evaluation derives from comparing an outcome against a reference point parameter and by weighting any discrepancy by an uncertainty parameter. Maladaptive evaluation is proposed to occur when these parameters do not reflect the true context statistics. Depending on which parameter is altered, different forms of maladaptive evaluation emerge, each associated with specific clinical conditions. This model highlights how the concept of reference-dependent evaluation can elucidate several clinical conditions, including perfectionism, depression, and addiction.

People constantly assign a value to stimuli and outcomes around them. These *evaluation processes* underlie "hot" aspects of cognition such as motivation, emotion, and affect. In a variety of psychopathologies, impairments in evaluation processes are critical. For example, pathologically anxious individuals (e.g., arachnophobic patients) attribute an excessive negative value to certain stimuli (e.g., spiders; Lang et al., 1998). People experiencing depression evaluate a multiplicity of aspects of their life as gloomy (Smith, 2013). Addiction leads to an overwhelming desire for consuming drugs (Robinson & Berridge, 2001). Hence, understanding how impairments in evaluation might contribute to psychopathology is paramount for research. Inspired by seminal studies in prospect theory (Kahneman & Tversky, 1979), contemporary models of evaluation highlight its reference-dependent nature (Koszegi &

Rabin, 2006; Louie et al., 2013, 2015; Rigoli, 2019; Rigoli, Friston, et al., 2016; N. Stewart, 2009; N. Stewart et al., 2006; Woodford, 2012): When attributing a value to an outcome, the brain automatically assesses the outcome not in isolation but relative to its context. For example, consider an individual who is purchasing a house and who discovers that the price of the house is £10 more than expected. Compare this with someone paying for a coffee and realizing that the price is £10 more than expected. Although objectively both individuals experience an equivalent unforeseen extra cost of £10, we would expect the second person to be way more upset than the first. This example stresses the idea that evaluation is

**Corresponding Author:**
Francesco Rigoli, Department of Psychology, City, University of London
E-mail: francesco.rigoli@city.ac.uk

reference-dependent—the subjective value of outcomes strongly depends on the context in which these outcomes are experienced (analogous phenomena have been reported in the social domain in monkeys; Brosnan & De Waal, 2003; this hints to the fact that, in evolutionary terms, reference dependency might be the product of a tendency for social comparison expressed by animals living in complex social groups).

Given the central role of evaluation in psychopathology and the central role of reference effects in evaluation, exploring reference effects in the context of psychopathology appears as a promising research avenue. What are the implications of reference-dependent models for the development and maintenance of psychopathology? Can reference-dependent models shed light on important psychopathological processes? To date, these questions remain open. In this article, we aim to develop a theoretical framework that applies reference-dependent models of evaluation to psychopathology. First, following recent literature, we describe a general computational model of evaluation (Rigoli, 2019; Woodford, 2012); we refer to this as the reference-dependent model of psychopathology (RDMP). Next, adopting this as framework, we examine how aberrant evaluation could arise in psychopathology. At the same time, the link between our approach and key concepts in clinical literature are highlighted, allowing us to shed new light on these concepts.

## The Model

In contemporary literature, three perspectives debate on how reference-dependent evaluation works. First, divisive normalization models propose that reference-dependent evaluation emerges because reward sensitivity decreases when previous outcomes are highly valuable (Louie et al., 2013, 2015). Second, decision by sampling postulates that previous experiences are sampled from memory and compared with the current outcome, resulting in reference-dependent evaluation (N. Stewart, 2009; N. Stewart et al., 2006). Third, expectation-as-reference accounts propose that reference-dependent evaluation arises because previous experiences set expectations to which outcomes are compared (Koszegi & Rabin, 2006; Rigoli, 2019; Rigoli, Friston, et al., 2016; Woodford, 2012); the RDMP relies on this latter approach. However, we do not aim at comparing the three approaches (for a detailed overview and comparison, see Rigoli, 2019) but at highlighting their similarity: All share the key principle that outcomes are evaluated considering the environmental statistics, with analogous implications for mental illness.

To introduce the RDMP, consider an environment in which a set of stimuli (e.g., money) can be experienced, each associated with a raw value (e.g., a monetary amount). For each stimulus, the calculation of the subjective value $V_R$ associated with the raw value $R$ depends on the following logistic function:

$$V_R = \frac{1}{1 + e^{-\frac{R - \mu}{\sigma}}} \tag{1}$$

This prescribes that the subjective value of a stimulus is $0 < V_R < 1$. The parameters $\mu$ and $\sigma$ (being $\sigma > 0$) are the reference point and the uncertainty, respectively (Fig. 1). These parameters capture the reference-dependent nature of evaluation: The subjective value, which is experienced at a subjective level and drives behavior, is not equivalent to the raw value, but it depends on some reference information. The RDMP proposes that subjective value can be experienced as either reward or punishment, occurring when $V_R > 0.5$ and $V_R < 0.5$, respectively (a neutral experience occurs when $V_R = 0.5$). According to this definition, reward is experienced when $R > \mu$, and punishment is experienced when $R < \mu$. Therefore, the reference point can be interpreted as the standard (or expectation) to which outcomes are compared and are evaluated as reward (i.e., better than the standard) or as punishment (i.e., worse than the standard; Fig. 1a). For example, the reference point $\mu$ might indicate the standard mark at school, implying that a better mark will be perceived as success and a worse mark as failure. The parameter $\sigma$ can be interpreted as the level of uncertainty about the own standard, prescribing how much a discrepancy from the reference point will be weighted (Fig. 1b). In other words, it determines how subjectively good or bad an outcome is when compared with the reference point. For example, if one has received a mark above/below the standard, the uncertainty parameter determines how subjectively good/bad the mark is. If there is high uncertainty, then a discrepancy will not be weighted much, minimizing the subjective distance from the reference point (Fig. 1b). Hence, the mark above/below the standard will not be considered too good/bad. Conversely, if there is low uncertainty, a discrepancy will be weighted heavily, maximizing the subjective distance from the reference point (Fig. 1b). Hence, the mark above/below the standard will be considered as very good/bad.

In short, thanks to the reference point $\mu$ and the uncertainty parameter $\sigma$, the RDMP highlights the reference-dependent nature of subjective value. Below, we explore how critical characteristics of psychopathology can be interpreted within the framework.
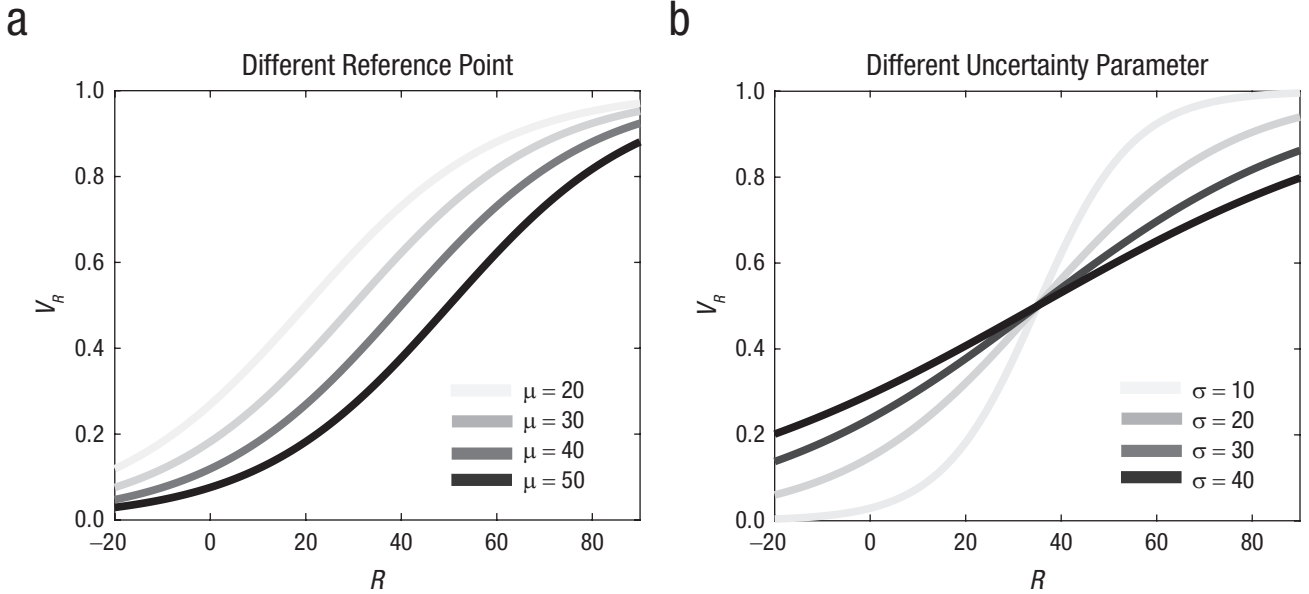
a

### Different Reference Point



b

### Different Uncertainty Parameter



**Fig. 1.** Role played by the model parameters. (a) Subjective value is graphed as a function of raw value for different values of reference point μ (σ = 20 for all lines). (b) Subjective value is graphed as a function of raw value for different values of uncertainty parameter σ (μ = 35 for all lines).

## Implications for Psychopathology

Consider an example of a context in which an agent can experience seven possible raw values (10, 20, 30, 40, 50, 60, and 70), each with equal probability. The RDMP model suggests that within this context, adaptive evaluation (Fig. 2a) occurs if the reference point μ corresponds to the contextual average (40 in this example) and the uncertainty parameter σ corresponds to the contextual standard deviation (20 in this example; Rigoli, 2019). In other words, adaptive evaluation occurs when an agent has a realistic representation of the context and its statistics and uses this representation to evaluate each stimulus appropriately relative to the others. Applying Equation 1 with μ = 40 and σ = 20 (reflecting the true context statistics), the subjective value of the different raw values corresponds to 0.1824, 0.2689, 0.3775, 0.5, 0.6225, 0.7311, and 0.8176, respectively (Table 1). Conversely, when the reference point μ and the uncertainty parameter σ do not reflect the true context statistics, evaluation is considered as maladaptive by the RDMP.

We propose that this maladaptive evaluation might explain important features of psychopathology. Below, we consider the two forms of maladaptive evaluation ensuing from the RDMP model (associated with altered reference point μ and altered uncertainty parameter σ, respectively), and we examine them in the context of psychopathology. Next, we analyze the processes leading to alterations of the model parameters, again exploring their link with psychopathology.

## *Reference point*

According to the RDMP, a case of maladaptive evaluation derives from a reference point μ that does not reflect the average of the contextual distribution but is higher or lower than the average. We first explore the case in which the reference point is higher than the contextual average. The example above (describing a context characterized by raw values 10, 20, 30, 40, 50, 60, and 70) can be examined, but now adopting a reference point μ that is higher than the average (in our example, higher than 40; the uncertainty parameter σ is still equal to 20). We consider two different instances because each can shed light on different aspects of psychopathology. First, the reference point μ might be high but not extreme (inside the range of the contextual distribution); in our example, it is equal to 70. Second, the reference point μ might be extremely high (outside the range of the contextual distribution), in our example equal to 100. Below, we explore each case and its relevance for psychopathology.

We first consider the case in which the reference point μ is equal to 70 (i.e., high but not extreme; see Fig. 2b and Tables 1 and 2). When comparing the results for μ = 40 and μ = 70, two key differences emerge. First, all subjective values are lower when μ = 70 (Table 1). In other words, a high reference point will lead to considering all possible outcomes as more negative. The second aspect concerns the distance in subjective value among outcomes that are adjacent in the distribution (e.g., 20 minus 10, or 30 minus 20, or 40 minus 30, etc.;
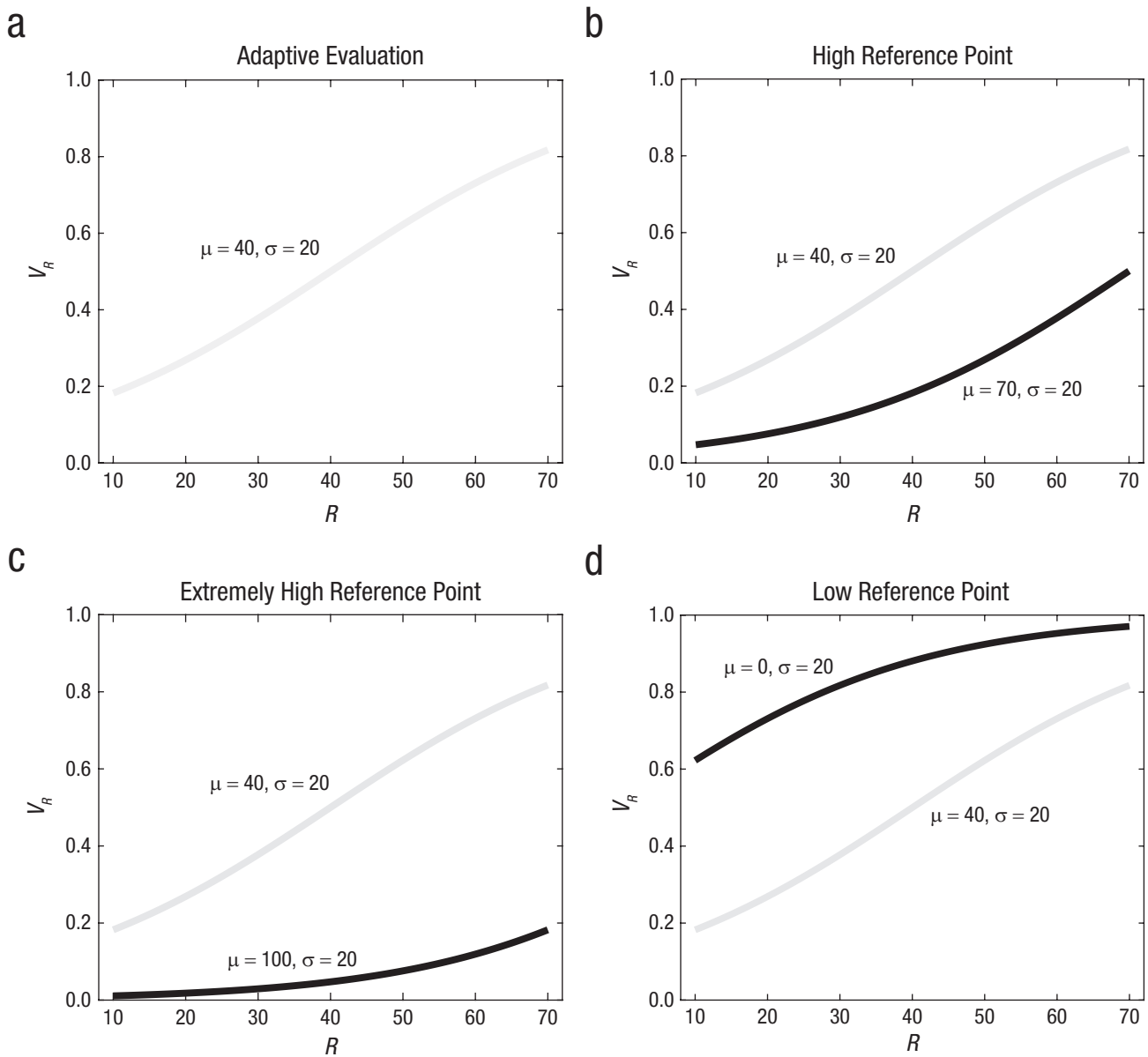
a

### Adaptive Evaluation



$\mu = 40, \sigma = 20$

b

### High Reference Point



$\mu = 40, \sigma = 20$

$\mu = 70, \sigma = 20$

c

### Extremely High Reference Point



$\mu = 40, \sigma = 20$

$\mu = 100, \sigma = 20$

d

### Low Reference Point



$\mu = 0, \sigma = 20$

$\mu = 40, \sigma = 20$

**Fig. 2.** Subjective value as a function of raw value for different parameter sets.

Table 2). When the reference point μ is equal to the contextual average (in our example, μ = 40), this distance is maximal near the average (in our example, it is maximal for 40 minus 30 and for 50 minus 40; Table 1). Conversely, when the reference point μ is high (in our example, μ = 70), this distance is maximal for a region above the average (in our example, it is maximal for 70 minus 60; Table 1). Moreover, in the lower tail of the contextual distribution, distances among adjacent outcomes are larger when the reference point is close to the contextual average compared with when it is high. In our example, for 20 minus 10, 30 minus 20, 40 minus 30, and 50 minus 40, the distance in subjective

value is larger when μ = 40 compared with when μ = 70. In short, when comparing a reference point close to the contextual average with a high reference point, the distance for adjacent outcomes is larger, except for a region at the high end of the distribution.

We argue that this scenario might help explaining important phenomena in psychopathology. Several clinical conditions, such as eating disorders, obsessive compulsive disorders, social anxiety, and depression, are associated with both perfectionism and dichotomous thinking (Burns & Fedewa, 2005; Byrne et al., 2004; Egan et al., 2007, 2011; Flett & Hewitt, 2002). Moreover, perfectionism and dichotomous thinking are

**Table 1.** Subjective Value ($V_R$) of Outcomes for Different Parameter Sets

| Outcome (raw value) | μ = 40, σ = 20 | μ = 70, σ = 20 | μ = 100, σ = 20 | μ = 0, σ = 20 | μ = 40, σ = 10 | μ = 40, σ = 40 |
|---|---|---|---|---|---|---|
| 10 | 0.1824 | 0.0474 | 0.0110 | 0.6225 | 0.0474 | 0.3208 |
| 20 | 0.2689 | 0.0759 | 0.0180 | 0.7311 | 0.1192 | 0.3775 |
| 30 | 0.3775 | 0.1192 | 0.0293 | 0.8176 | 0.2689 | 0.4378 |
| 40 | 0.5000 | 0.1824 | 0.0474 | 0.8808 | 0.5000 | 0.5000 |
| 50 | 0.6225 | 0.2689 | 0.0759 | 0.9241 | 0.7311 | 0.5622 |
| 60 | 0.7311 | 0.3775 | 0.1192 | 0.9526 | 0.8808 | 0.6225 |
| 70 | 0.8176 | 0.5000 | 0.1824 | 0.9707 | 0.9526 | 0.6792 |

themselves correlated (Burns & Fedewa, 2005; Egan et al., 2007). Perfectionism occurs when all possible outcomes are considered as negative except for only those at the very high end of the distribution (Egan et al., 2007, 2011; Flett & Hewitt, 2002). Moreover, even these outcomes at the top are usually experienced not with positive feelings but with just a sense of relief. For example, the top mark at school might not be perceived as a great achievement to be celebrated but simply as the minimum to be expected. This picture of perfectionism fits with the scenario described by the RDMP in which a high reference point is implemented. In our example, all outcomes are evaluated as negative (i.e., they have a subjective value smaller than 0.5) except for the outcome of 70, which is associated with a neutral value (equal to 0.5). This scenario captures the notion that in perfectionism expectations (captured by the reference point μ) are too high, resulting in disappointment (when the outcome is worse than expectations) or, at best, in relief (when the outcome matches expectations).

Dichotomous thinking occurs when possible outcomes are grouped in two opposing categories (Byrne et al., 2004). Moreover, dichotomous thinking is characterized by polarization; it maximizes the perceived distance among categories of outcomes and minimizes the distance within each category (Byrne et al., 2004). This picture of dichotomous thinking also fits with the scenario described by the RDMP in which a high reference point is implemented. In our example, comparing the condition in which μ = 70 with the condition in which μ = 40, the difference in subjective value is minimized for 10, 20, 30, 40, 50, and 60; in other words, these outcomes are perceived as more similar (Table 2). Hence, the RDMP predicts that a high reference point will group these outcomes together. At the same time, comparing the condition in which μ = 70 with the condition in which μ = 40, the distance between 60 and 70 is enhanced: These outcomes are perceived as more far apart (Table 2). Therefore, in line with the notion of dichotomous thinking, the RDMP predicts that a high

reference point groups the outcomes of 10, 20, 30, 40, 50, and 60 together while treating the outcome of 70 as a separate category. Altogether, when a high reference point is implemented in the RDMP, perfectionism and dichotomous thinking both emerge. Thus, the RDMP offers a formal description of perfectionism and dichotomous thinking and explains why these are commonly observed together empirically (Burns & Fedewa, 2005; Egan et al., 2007).

The scenario of a high reference point might shed light also on addiction. Similar to the RDMP, influential models of addiction (Koob, 1999; Koob & Le Moal, 2001, 2008; Sterling, 2012) maintain that subjective value derives from comparing an outcome with a set point (although they do not postulate any uncertainty parameter). According to these models, repeated drug consumption would overincrease the set point, leading to both tolerance and withdrawal: The same drug dose would lead to a decreased subjective value (tolerance), and abstinence would lead to a very unpleased state (withdrawal). Moreover, according to set-point models (Koob, 1999; Koob & Le Moal, 2001, 2008; Sterling, 2012), an overincrease in set point would explain why, once addiction has established, natural incentives such as food or social interactions lose their appeal: They would pale in comparison with the newly established set point. A similar argument ensues from the RDMP. In our example, imagine that drug consumption is associated with a raw value of 70, no drug consumption with a raw value of 30, and the reference point is μ = 40. We can also consider natural incentives such as food and social interaction as associated with raw values of 40 and 50, respectively. In this context, drug consumption would lead to a highly pleasurable experience ($V_R = 0.8176$) and no drug consumption to a slightly negative experience ($V_R = 0.3775$; Table 1). Food and social interaction would be experienced as neutral ($V_R = 0.5$) and pleasant ($V_R = 0.6225$), respectively. Repeated drug use, associated with an increased frequency of excessively pleasurable experiences, would eventually boost the reference point to μ = 70, resulting in addiction.

**Table 2.** Difference in Subjective Value ($V_R$) Between Adjacent Outcomes for Different Parameter Sets

| Adjacent outcome (raw value) | μ = 40, σ = 20 | μ = 70, σ = 20 | μ = 100, σ = 20 | μ = 0, σ = 20 | μ = 40, σ = 10 | μ = 40, σ = 40 |
|---|---|---|---|---|---|---|
| 20 – 10 | 0.0865 | 0.0285 | 0.0070 | 0.1086 | 0.0718 | 0.0567 |
| 30 – 20 | 0.1086 | 0.0433 | 0.0113 | 0.0865 | 0.1497 | 0.0603 |
| 40 – 30 | 0.1225 | 0.0632 | 0.0181 | 0.0632 | 0.2311 | 0.0622 |
| 50 – 40 | 0.1225 | 0.0865 | 0.0285 | 0.0433 | 0.2311 | 0.0622 |
| 60 – 50 | 0.1086 | 0.1086 | 0.0433 | 0.0285 | 0.1497 | 0.0603 |
| 70 – 60 | 0.0865 | 0.1225 | 0.0632 | 0.0181 | 0.0718 | 0.0567 |

This would imply that consuming the same drug dose becomes no more pleasurable ($V_R$ = 0.5; analogous to tolerance) and that not consuming any drug becomes highly unpleasant ($V_R$ = 0.1192; analogous to withdrawal). Food ($V_R$ = 0.1824) and social interaction ($V_R$ = 0.2689) would now both become unpleasant. These are analogous to predictions of set-points models (Koob, 1999; Koob & Le Moal, 2001, 2008; Sterling, 2012).

However, the RDMP adds another aspect not considered by standard set-point models: It predicts that a reference point increase (caused by repeated drug consumption) will affect the distances among outcomes (Table 2). According to the RDMP, before addiction (i.e., when μ = 40), the distance between drug consumption and social interaction is 0.8176 – 0.6225 = 0.1951, whereas once addiction has established (i.e., when μ = 70), it increases to 0.5 – 0.2689 = 0.2311. In other words, the RDMP predicts that once addiction has developed, the value of drug consumption compared with natural incentives (e.g., social interaction) will be magnified; set-point models do not make such prediction. Moreover, before addiction (i.e., when μ = 40), the distance between food and social interaction is 0.1225, whereas once addiction has established (i.e., when μ = 70), it decreases to 0.0865. Likewise, before addiction (i.e., when μ = 40), the distance between food and no drug consumption is 0.1225, whereas once addiction has established (i.e., when μ = 70), it decreases to 0.0632. In other words, the RDMP predicts that once addiction has developed, different natural incentives will all be perceived as rather similar, leading to an indiscriminate affective response: The ability to discriminate among outcomes, and to choose accordingly, will be impaired. This RDMP prediction, absent in set-point theories, fits with empirical observations showing that individuals with addiction exhibit a similar level of disinterest for incentives besides the drug (Hatzigiakoumis et al., 2011; Leventhal et al., 2014).

We now consider the case in which the reference point μ is extremely high (outside the range of the contextual distribution), in our example equal to 100 (see Fig. 2c and Tables 1 and 2). Here all subjective values are experienced as punishments (i.e., they have subjective value lower than 0.5) and are more negative compared with when μ = 40 or μ = 70 (Table 1). Moreover, the distance in subjective value among adjacent outcomes is very small (the largest being equal to 0.0433; Table 2), implying that different outcomes are perceived as rather similar, leading to an indiscriminate (and highly negative) affective response to all of them. We argue that this scenario is appropriate to describe the affective processes characterizing depression. In line with this scenario, people suffering from depression feel severely negative affect independent of the outcome they experience (Rottenberg, 2005). Even objectively positive outcomes do not alter the negative affect felt by these people (anhedonia; Pizzagalli, 2014). Moreover, several studies have reported that depression, especially when severe, is characterized by decreased emotional discrimination (i.e., an emotional response expressed both at the behavioral and physiological levels that is similar for both salient and neutral stimuli; Bylsma et al., 2008; Rottenberg, 2017; Rottenberg et al., 2005; Rottenberg & Johnson, 2007). This observation fits with the scenario described by the RDMP in which the reference point μ is extremely high: As we have just demonstrated, in this scenario different outcomes elicit a very similar affective response despite the outcomes being objectively very different. Note that the two characteristics of depression that (a) outcomes elicit negative affect and (b) outcomes elicit indiscriminate emotional responses both ensue naturally from the RDMP, without any further assumption.

Altogether, a large reference point μ is postulated in perfectionism (combined with dichotomous thinking), addiction, and (in the extreme case) depression. How does the RDMP integrate these concepts? Perfectionism and addiction are conceptually analogous: Both imply disdain of virtually all stimuli except one—the very top outcome and the addictive substance for perfectionism

and addiction, respectively. However, the fact that these stimuli (the very top outcome and the addictive substance) are unrelated implies that perfectionism and addiction are ultimately very different conditions. Regarding depression, the RDMP associates a higher reference point to this pathology compared with perfectionism and addiction: During depression, nothing (not even the top outcome or the substance) appears as positive or even neutral. However, some features of perfectionism and addiction as depicted by the RDMP (e.g., the fact that most events are experienced as negative) fit with the diagnosis of depression found in clinical manuals (American Psychiatric Association, 2013). In other words, within the RDMP, the concept of depression is, strictly speaking, different from perfectionism and addiction; however, perfectionism and addiction exhibit features consistent with the standard diagnosis of depression. This is in line with observations of a comorbidity between depression and addiction (Grant & Harford, 1995; evidence indicates that depression can ensue from addiction and often disappears after detoxification, Brown & Schuckit, 1988) and of a link between depression and perfectionism (Flett & Hewitt, 2002).

Finally, according to the RDMP, maladaptive evaluation occurs when the reference point $\mu$ is lower than the average of the contextual distribution, in our example when the reference point is equal to 0 (see Fig. 2d and Tables 1 and 2). Here, the pattern is symmetrical to when the reference point is larger than the average. Now, all subjective values are experienced as reward (i.e., they have subjective value higher than 0.5) and are more positive compared with when $\mu = 40$, $\mu = 70$, or $\mu = 100$ (Table 1). Moreover, in comparison with when $\mu = 40$ (i.e., equal to the contextual distribution), the distance in subjective value among adjacent outcomes is smaller, implying that different outcomes are perceived as more similar, leading to a less discriminate affective response (Table 2). Regarding mental health, this scenario fits with the affective experience observed in mania. In line with this scenario, empirical evidence shows that patients with mania exhibit more positive emotional responses when exposed to negative, neutral, or positive stimuli (Gruber et al., 2008; Johnson et al., 2005; Lennox et al., 2004; Meyer et al., 2001). Whether their ability to discriminate among outcomes is also impaired, as predicted by this scenario, remains to be explored empirically.

In conclusion, the RDMP proposes that alterations of the reference point, either upward or downward compared with the contextual average, lead to maladaptive evaluation. Conditions of altered reference point as described by the RDMP have remarkable analogies with a variety of clinical conditions characterized by abnormal affective processes, including perfectionism, dichotomous thinking, addiction, depression, and mania. These analogies suggest that the RDMP might contribute to explain affective impairments observed in these clinical conditions. In essence, according to the RDMP, changes in the reference point have two implications: (a) They determine whether outcomes are perceived as more or less positive, and (b) they impair the ability to discriminate among outcomes. We stress that this picture goes beyond standard theories (e.g., the set-point model of addiction; Koob, 1999; Koob & Le Moal, 2001, 2008; Sterling, 2012) that focus exclusively on one of the implications of the RDMP (e.g., the set-point model explains only whether outcomes are perceived as more or less positive but not why discriminating abilities are impaired). Below, we explore the implications of the RDMP for psychopathology in the case of an altered uncertainty parameter $\sigma$.

## Uncertainty parameter

Even when the reference point describes the contextual average accurately, according to the RDMP maladaptive evaluation occurs when the uncertainty parameter $\sigma$ is altered (i.e., it is larger or smaller than the contextual standard deviation). We first focus on an uncertainty parameter $\sigma$ smaller than the contextual standard deviation. Consider the example above (describing a context characterized by raw values 10, 20, 30, 40, 50, 60, and 70; note that the actual $SD = 20$), but now with an uncertainty parameter $\sigma$ equal to 10 (see Fig. 3a and Tables 1 and 2; note that the reference point $\mu$ is still equal to 40, correctly reflecting the contextual average). Comparing the case in which $\sigma = 10$ and the case in which $\sigma = 20$, every time a reward (i.e., when $V_R > 0.5$) is experienced in one case, it is also experienced in the other case; likewise, every time a punishment (i.e., when $V_R < 0.5$) is experienced in one case, it is also experienced in the other case (Table 1). In addition, contrary to cases in which the reference point $\mu$ changes (see above), on average subjective values for the outcomes are equivalent in these two cases. However, comparing the case in which $\sigma = 10$ with the case in which $\sigma = 20$, the distance in subjective value among adjacent outcomes is magnified near the contextual average (40), and it is minimized far from the average (i.e., for 20 − 10 and for 70 − 60; Table 2). The case described by the RDMP in which $\sigma$ is excessively small has analogies with clinical conditions characterized by exaggerated emotional sensitivity for both negative and positive stimuli, such as in borderline personality disorder. An influential proposal (Herpertz et al., 1997; Linehan, 1993; Lynch et al., 2006) is that patients suffering
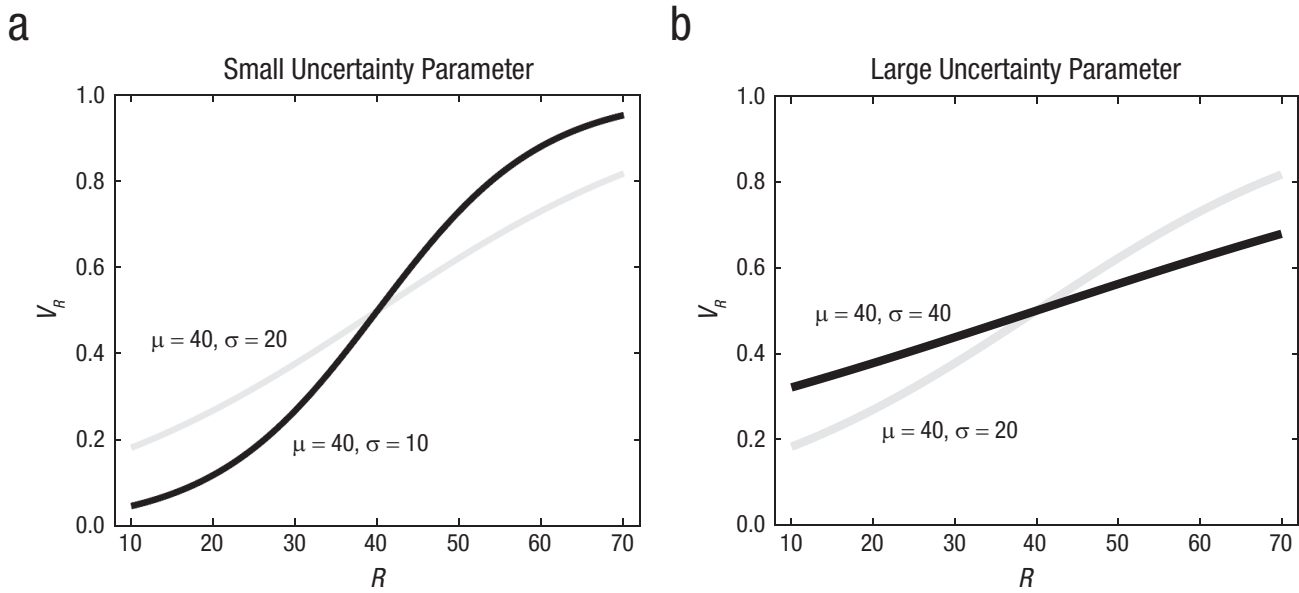
a

### Small Uncertainty Parameter



b

### Large Uncertainty Parameter



**Fig. 3.** Subjective value as a function of raw value for different parameter sets.

from this disorder exhibit intense affective reactions even when exposed to mild events (although empirical testing of this has produced mix findings; see Rosenthal et al., 2008). According to this proposal, these patients would experience exaggerated negative emotions in response to mildly negative stimuli as well as exaggerated overenthusiasm when something mildly positive occurs. This idea is consistent with the scenario offered by the RDMP when the uncertainty parameter σ is small.

We now examine the case in which the uncertainty parameter is higher than the contextual standard deviation, in our example σ = 40 (see Fig. 3b and Tables 1 and 2). Comparing the case in which σ = 40 with the case in which σ = 20 (i.e., reflecting the actual standard deviation), every time a reward (i.e., when $V_R > 0.5$) is experienced in one case, it is also experienced in the other case; likewise, every time a punishment (i.e., when $V_R < 0.5$) is experienced in one case, it is also experienced in the other case (Table 1). In addition, contrary to cases in which the reference point μ changes (see above), on average subjective values for the outcomes are equivalent in these two cases. However, comparing the case in which σ = 40 with the case in which σ = 20, the distance in subjective value among adjacent outcomes is much smaller (the largest being equal to 0.0632; Table 2). In other words, when the uncertainty parameter is large, affective experiences are predicted to be attenuated, or very close to a neutral evaluation (i.e., when $V_R = 0.5$, associated with $R = μ$). Regarding psychopathology, this scenario fits with the concept of apathy (Marin, 1991, 1996; Starkstein & Leentjens, 2008). Apathy, defined as lack of motivation

associated with numbed emotional responses, is common in a variety of disorders, including schizophrenia, dementia, Parkinson's disease, and damage of the frontal lobes (Marin, 1996). Note that this account of apathy excludes conditions, such as depression or anxiety, in which inactivity results from strong avoidance motivation and from intense emotions (e.g., depressed patients may staunchly resist treatment or socialization; Levy et al., 1998; Marin, 1996). Inactivity exhibited by people affected by apathy is passive and compliant, reflecting poor motivation and numbed emotion. The RDMP offers a formal interpretation of apathy as arising from a large uncertainty parameter σ, leading to perception of all outcomes as similar and as rather neutral. This inability to discriminate among outcomes would explain why motivation is poor and emotions are numbed.

In short, the RDMP proposes that maladaptive evaluation might derive not only from an altered reference point but also from an altered uncertainty parameter. This allows the model to account for conditions, such as excessive emotional sensitivity in borderline patients and apathy, that cannot be explained by an altered reference point. Below, we explore the processes leading to alterations in the model parameters and to maladaptive evaluation.

## *Etiology of maladaptive evaluation*

A fundamental research question is where the reference point and the uncertainty parameter come from and why these sometimes become altered. This fundamental

question remains largely open because research on how context parameters for evaluation develop is in its infancy (Rigoli et al., 2018). However, it is helpful to discuss some general processes that are likely to be at play, focusing on four factors: genetic influences, direct learning, social influence, and drug-related effects.

First, genetic factors are arguably critical in affecting the parameters of the model underlying evaluation. For example, the brain is hardwired to perceive some stimuli as painful, and pain is normally perceived as punishment. These innate constraints are arguably critical in shaping the model parameters. Genetic factors are also likely to contribute to individual differences in these parameters, for example as reflected in child temperament (Saudino, 2005). In other words, a propensity for entertaining lower or higher reference points or lower or higher uncertainty parameters might emerge at least partially as a genetic trait.

Second, parameters derive, at least partially, from direct experience with salient stimuli in different environments or contexts. Substantial evidence has shown that experience in childhood and adolescence is critical in determining an adult's predisposition to mental illness (e.g., Aas et al., 2016; Mandelli et al., 2015). Within the RDMP, one can attempt to explain this by relying on the notion that the model parameters reflect the statistics of the environment experienced early in life, which are then translated to adulthood largely unchanged. Hence, early exposure to an unhealthy environment would develop altered parameters that are at play also in adulthood. However, this explanation fits poorly with empirical data showing that early traumatic events predispose adults to depression (Mandelli et al., 2015): This explanation predicts that traumatic events should lead to an excessively low reference point μ, whereas depression would be characterized by an excessively high reference point μ (see above). However, an alternative explanation also arises from the RDMP. This explanation argues that early life experience is not so much important in affecting the model parameters but, rather, in establishing the learning rate (which determines how the parameters change with novel experience). In human experience, the environment changes continuously, and tracking change requires updating the model parameters with an appropriate learning rate (a detailed account of how the learning rate develops will not be pursued here but is left for future work—see previous literature for some insight on this; Behrens et al., 2007; Browning et al., 2015). An excessively large or small learning rate results in altered parameters. An excessively small learning rate might result in an inability to flexibly tune to the environment. This might characterize forms of depression following bereavement (Clayton, 1990), during which a person is unable to "learn" that the environment is now worse than before

and to reset subjective values accordingly, resulting in depression. An excessively large learning rate might result in parameters that are unstable even when the environment does not change, leading to dramatic mood swings, such as in bipolar disorder (where an extremely high reference point, conducive of depression, rapidly alternates with an extremely low reference point, conducive of mania; Gottschalk et al., 1995).

Early life might be particularly important in establishing the learning rate, and early traumatic experience might be conducive of alterations of the learning rate (genetic factors might also be important). It is possible that whether early traumatic events produce excessively high or low learning rates (predisposing for bipolar disorder and depression, respectively) depends on the temporal pattern of these traumatic events. Experiencing traumatic events all close in time would lead to the interpretation that the reference point μ can change abruptly, leading to a large learning rate and to a predisposition for bipolar disorder. Conversely, experiencing the same traumatic events but now sparse in time would lead to the interpretation that the reference point μ is rather fixed, leading to a small learning rate and to a predisposition for depression. This possibility fits with empirical observations showing that early trauma can predispose to either depression (Mandelli et al., 2015) or bipolar disorder (Aas et al., 2016). However, the influence of the traumatic temporal pattern advocated here is a novel aspect that remains to be investigated empirically.

Moreover, traumas are associated not only with depression and bipolar disorder but also with other psychopathologies such as addiction (S. H. Stewart, 1996) and borderline personality (Ball & Links, 2009). Learning rate alterations caused by traumas might be a factor in these conditions, too. Regarding addiction, an excessively high learning rate might facilitate the development of this disorder (explaining the link between addiction and bipolar disorder; Altamura, 2007), whereas an excessively low learning rate might interfere with recovery once addiction has established (explaining the link between addiction and depression; Grant & Harford, 1995). Regarding borderline personality, traumatic experiences might also alter the learning rate, although not about the reference point but about the uncertainty parameter. This would result in an inability to adjust the uncertainty parameter when the contextual standard deviation changes, hence resulting in an abnormal uncertainty parameter (as examined above).

Arguably, parameters depend not only on direct experience with outcomes but also on social influence. For example, parents might teach their children that the top mark at school is the norm (Cook & Kearney, 2009, 2014). This would lead their children to develop an

exaggerated reference point μ, facilitating perfectionism and dichotomous thinking or even depression. Furthermore, different parents might express their opinions about standards with varying degrees of confidence: Some parents might be absolutely sure about their standards, contrary to other parents expressing their standards cautiously. Opinions expressed with high confidence would lead children to develop high confidence in their standards, too, reflected in small values of σ. Conversely, parents' opinions expressed with low confidence would lead children to develop uncertainty about their standards, reflected in large values of σ.

Finally, certain drugs might affect the model parameters. Large evidence indicates that the neurotransmitter dopamine is central in evaluation (Wise, 2004). Consistent with a role in reference-dependent evaluation, dopaminergic responses are tuned to expectations (Dabney et al., 2020; Schultz et al., 1997; Tobler et al., 2005). Moreover, an association between behavioral reference effects and dopaminergic tuning has been reported (Rigoli, Friston, & Dolan, 2016; Rigoli, Rutledge, et al., 2016). This suggests that within the RDMP, drugs affecting the dopaminergic system (including addictive drugs) might be interpreted as altering the reference point parameter. Serotonin is another neurotransmitter important in evaluation, and serotoninergic drugs are well-established treatments for depression (Hieronymus et al., 2018). Within the RDMP, the effect of these drugs might be interpreted as increasing the learning rate so that, after some new learning occurs, the reference point would decrease and adapt to the ongoing environment. This implies that, once the learning rate is boosted by serotoninergic drugs, new learning is required for change to occur, in line with observations that serotoninergic drugs have delayed action (Michely et al., 2020).

In short, although the precise mechanisms responsible for the development of the model parameters remain to be understood, four general factors can be advocated, comprising genetic influences, direct learning, social influence, and drug-related effects.

## Discussion

Evaluation processes are critical in psychopathology, and reference effects are critical in evaluation (Louie et al., 2013, 2015; Koszegi & Rabin, 2006; Rigoli, 2019; Rigoli, Friston, et al., 2016; N. Stewart, 2009; N. Stewart et al., 2006; Woodford, 2012). Yet reference-dependent evaluation in psychopathology remains to be investigated. In this article, we attempt to fill this gap by offering a computational model of evaluation processes in psychopathology grounded on reference dependency. The advantage of this approach is threefold.

First, it offers a mechanistic description of how evaluation abilities might go awry and lead to mental illness. Second, it proposes a formal interpretation of standard concepts in the literature such as perfectionism, dichotomous thinking, and apathy. To date, these concepts have been usually described verbally; our model recasts these in computational terms, contributing to clearer and less ambiguous definitions. Third, the model pinpoints to the precise nature of the connections among concepts, for example between perfectionism and dichotomous thinking. So far, research has reported these connections at an empirical level, yet the nature of these connections has remained unclear at a theoretical level; the model offers a formal description of this nature.

Here we have examined the RDMP relative to a variety of phenomena across different clinical diagnoses. This raises the question of which phenomena can be fruitfully explored by the model. In other words, what is the potential scope of the RDMP? Contemporary theories maintain that reference dependency is at the core of evaluation, implying that the RDMP can potentially contribute to understand clinical conditions in which evaluation has gone awry, such as conditions characterized by abnormal motivation, emotion, and affect; a promising avenue is to apply the model to these. The RDMP appears as less relevant for clinical aspects in which evaluative processes are peripheral, such as hallucinations, learning deficit, or dementia, to mention a few (although the model can still be useful to examine secondary affective processes ensuing from these aspects).

In presenting the RDMP, an implicit assumption has been that an individual represents a unique environment, described by one reference point and one uncertainty parameter. However, representations of different contexts (each with its own parameter set) might coexist in the brain, and the ongoing context might determine which representation is currently engaged. For example, an individual might represent the family context as separated from the work context and activate the former at home and the latter in the workplace. An interesting research avenue is exploring the implications for psychopathology of having multiple context representations. Consider addiction: People affected by it might have a high reference point (proposed as underlying this disorder by the RDMP) only in some contexts (e.g., at a party) and not in others (e.g., at home). This implies drug craving in the first but not in the second context (i.e., at the party and not at home), allowing the RDMP to explain why some contexts can elicit drug craving even after prolonged abstinence: They would remain associated with a high reference point even when most other contexts are not anymore.

Potentially, this picture can contribute to integrating set-point models (Koob, 1999; Koob & Le Moal, 2001, 2008; Sterling, 2012) with incentive-sensitization models of addiction (which emphasize how specific contexts boost drug value; Robinson & Berridge, 2001, 2008). Moreover, this has implications for interpreting recovery from addiction. Given that craving can arise even after prolonged abstinence, some have argued that, once established, addiction is irreversible (Volkow et al., 2016). Conversely, the RDMP suggests that, in principle, addiction can be fully counteracted (Heather et al., 2018) in as much as new learning can lower the reference point. However, assuming existence of multiple context representations, recovery would be complete only when new learning (and the ensuing lowering of the reference point) occurs for each context. As long as learning has not occurred in a context, the RDMP predicts that craving will ensue during exposure to that context.

Although this article focuses on the evaluation of external stimuli, the model applies equally to the evaluation of interoceptive stimuli, such as visceral and cardiac signals. The correct evaluation of interoceptive stimuli is considered to be crucial for adaptive regulation, emotional processing, and conscious presence (Barrett, 2017; Craig, 2002; Iodice et al., 2019; Seth et al., 2012). Conversely, the failure to correctly process and evaluate interoceptive stimuli may have psychopathological consequences, such as the dysregulation of behavior observed in eating disorders or the aberrant interpretation of bodily stimuli observed in chronic pain (Barca & Pezzulo, 2020; Barrett et al., 2016; Khalsa et al., 2018; Paulus et al., 2019). Establishing to what extent these and other psychopathological conditions could be explained within the model proposed here is an open objective for future research.

Although evaluation is pivotal in processes such as emotion, feeling, and motivation, it is not the only aspect involved in these processes. For example, awareness about the own evaluation processes is likely to be critical (Lambie & Marcel, 2002). In other words, some individuals have poor understanding of why they evaluate outcomes the way they do, whereas other individuals have substantial insight on this. The role of awareness is arguably important in many respects. A possibility is that awareness might foster strategies to develop more "healthy" model parameters, hence promoting adaptive evaluation. An intriguing possibility is that some of the beneficial effects of psychotherapy might be interpreted in this way—as promoting awareness of the patient's evaluation process (and the associated model parameters) and as promoting the development of strategies to change this process (and hence the parameters; Papalini et al., 2020). Given this reasoning, a promising

research avenue is to explore the role of awareness about evaluation processes, the impact of awareness in regulating the model parameters, and the role of psychotherapy in promoting more healthy parameters via promoting awareness as well as via other strategies.

More generally, the RDMP can inspire process theories examining how treatments of mental illness work, thus contributing to improve these treatments. The effects of certain interventions can be interpreted, at least partially, as fostering more appropriate model parameters. For example, mindfulness has emerged as an effective intervention for a variety of clinical conditions (Sauer et al., 2013). Within our model, by promoting focus on and acceptance of the ongoing environment, mindfulness practices can be interpreted as strategies aimed at readapting the model parameters. An intriguing avenue is to adopt the RDMP as a framework for developing computational process theories of treatments (i.e., theories in which the processes are examined at a computational level) such as mindfulness.

Finally, we stress potential future developments of the RDMP. First, whereas here we pursue a transdiagnostic approach encompassing phenomena common to various conditions, the model could be applied to develop profiles of each disorder. Second, although we suggest an important role for learning in shaping the model parameters, here we do not formally examine it; a detailed analysis of learning in the context of the RDMP is a promising research avenue. Third, it is important to explore the neural processes underlying abnormal model parameters and learning. Research has explored the neural aspects of certain mental illnesses, on the one hand, and of reference-dependent evaluation, on the other; however, these two bodies of knowledge remain to be integrated. Fourth, as discussed above, the RDMP can inspire computational process theories of treatments.

In summary, we propose a computational model of evaluation in psychopathology grounded on reference dependency. This offers a formal description of key processes underlying mental illness and a precise interpretation of concepts (and of their links) developed by previous verbal theories.

## Transparency

## ORCID iD

Francesco Rigoli 🔟 https://orcid.org/0000-0003-2233-934X

## References

Aas, M., Henry, C., Andreassen, O. A., Bellivier, F., Melle, I., & Etain, B. (2016). The role of childhood trauma in bipolar disorders. *International Journal of Bipolar Disorders*, *4*(1), Article 2. https://doi.org/10.1186/s40345-015-0042-0

Altamura, A. C. (2007). Bipolar spectrum and drug addiction. *Journal of Affective Disorders*, *99*(1–3), 285. https://doi.org/10.1016/j.jad.2006.09.005

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Ball, J. S., & Links, P. S. (2009). Borderline personality disorder and childhood trauma: Evidence for a causal relationship. *Current Psychiatry Reports*, *11*(1), 63–68.

Barca, L., & Pezzulo, G. (2020). Keep your interoceptive streams under control: An active inference perspective on anorexia nervosa. *Cognitive, Affective, & Behavioral Neuroscience*, *20*(2), 427–440. https://doi.org/10.3758/s13415-020-00777-6

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23.

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), Article 20160011. https://doi.org/10.1098/rstb.2016.0011

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221.

Brosnan, S. F., & De Waal, F. B. (2003). Monkeys reject unequal pay. *Nature*, *425*(6955), 297–299.

Brown, S. A., & Schuckit, M. A. (1988). Changes in depression among abstinent alcoholics. *Journal of Studies on Alcohol*, *49*(5), 412–417.

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

Burns, L. R., & Fedewa, B. A. (2005). Cognitive styles: Links with perfectionistic thinking. *Personality and Individual Differences*, *38*(1), 103–113.

Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review*, *28*(4), 676–691.

Byrne, S. M., Cooper, Z., & Fairburn, C. G. (2004). Psychological predictors of weight regain in obesity. *Behaviour Research and Therapy*, *42*(11), 1341–1356.

Clayton, P. J. (1990). Bereavement and depression. *The Journal of Clinical Psychiatry*, *51*(Suppl.), 34–40.

Cook, L. C., & Kearney, C. A. (2009). Parent and youth perfectionism and internalizing psychopathology. *Personality and Individual Differences*, *46*(3), 325–330.

Cook, L. C., & Kearney, C. A. (2014). Parent perfectionism and psychopathology symptoms and child perfectionism. *Personality and Individual Differences*, *70*, 1–6.

Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*(8), 655–666.

Dabney, W., Kurth-Nelson, Z., Uchida, N., Starkweather, C. K., Hassabis, D., Munos, R., & Botvinick, M. (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature*, *577*(7792), 671–675.

Egan, S. J., Piek, J. P., Dyck, M. J., & Rees, C. S. (2007). The role of dichotomous thinking and rigidity in perfectionism. *Behaviour Research and Therapy*, *45*(8), 1813–1822.

Egan, S. J., Wade, T. D., & Shafran, R. (2011). Perfectionism as a transdiagnostic process: A clinical review. *Clinical Psychology Review*, *31*(2), 203–212.

Flett, G. L., & Hewitt, P. L. (2002). *Perfectionism: Theory, research, and treatment*. American Psychological Association.

Gottschalk, A., Bauer, M. S., & Whybrow, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Archives of General Psychiatry*, *52*(11), 947–959.

Grant, B. F., & Harford, T. C. (1995). Comorbidity between DSM-IV alcohol use disorders and major depression: Results of a national survey. *Drug and Alcohol Dependence*, *39*(3), 197–206.

Gruber, J., Johnson, S. L., Oveis, C., & Keltner, D. (2008). Risk for mania and positive emotional responding: Too much of a good thing? *Emotion*, *8*(1), 23–33. https://doi.org/10.1037/1528-3542.8.1.23

Hatzigiakoumis, D. S., Martinotti, G., Di Giannantonio, M., & Janiri, L. (2011). Anhedonia and substance dependence: Clinical correlates and treatment options. *Frontiers in Psychiatry*, *2*, Article 10. https://doi.org/10.3389/fpsyt.2011.00010

Heather, N., Best, D., Kawalek, A., Field, M., Lewis, M., Rotgers, F., & Heim, D. (2018). Challenging the brain disease model of addiction: European launch of the addiction theory network. *Addiction Research and Theory*, *26*(4), 249–255. https://doi.org/10.1080/16066359.2017.1399659

Herpertz, S., Gretzer, A., Steinmeyer, E. M., Muehlbauer, V., Schuerkens, A., & Sass, H. (1997). Affective instability and impulsivity in personality disorder: Results of an experimental study. *Journal of Affective Disorders*, *44*(1), 31–37.

Hieronymus, F., Lisinski, A., Nilsson, S., & Eriksson, E. (2018). Efficacy of selective serotonin reuptake inhibitors in the absence of side effects: A mega-analysis of citalopram and paroxetine in adult depression. *Molecular Psychiatry*, *23*(8), 1731–1736.

Iodice, P., Porciello, G., Bufalari, I., Barca, L., & Pezzulo, G. (2019). An interoceptive illusion of effort induced by false heart-rate feedback. *Proceedings of the National Academy of Sciences, USA*, *116*(28), 13897–13902.

Johnson, S. L., Ruggero, C. J., & Carver, C. S. (2005). Cognitive, behavioral, and affective responses to reward: Links with hypomanic symptoms. *Journal of Social and Clinical Psychology*, *24*(6), 894–906.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*, 263–292. https://doi.org/10.2307/1914185

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P., Simmons, W. K., Stein, M. B., . . . Interoception Summit 2016 participants. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(6), 501–513. https://doi.org/10.1016/j.bpsc.2017.12.004

Koob, G. F. (1999). The role of the striatopallidal and extended amygdala systems in drug addiction. *Annals of the New York Academy of Sciences*, *877*(1), 445–460.

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology*, *24*(2), 97–129.

Koob, G. F., & Le Moal, M. (2008). Addiction and the brain antireward system. *Annual Review of Psychology*, *59*, 29–53.

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165.

Lambie, J. A., & Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review*, *109*(2), 219–259. https://doi.org/10.1037/0033-295x.109.2.219

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, *44*(12), 1248–1263.

Lennox, B. R., Jacob, R., Calder, A. J., Lupson, V., & Bullmore, E. T. (2004). Behavioural and neurocognitive responses to sad facial affect are attenuated in patients with mania. *Psychological Medicine*, *34*(5), 795–802. https://doi.org/10.1017/s0033291704002557

Leventhal, A. M., Trujillo, M., Ameringer, K. J., Tidey, J. W., Sussman, S., & Kahler, C. W. (2014). Anhedonia and the relative reward value of drug and nondrug reinforcers in cigarette smokers. *Journal of Abnormal Psychology*, *123*(2), 375–386. https://doi.org/10.1037/a0036384.

Levy, M. L., Cummings, J. L., Fairbanks, L. A., Masterman, D., Miller, B. L., Craig, A. H., Paulsen, J. S., & Litvan, I. (1998). Apathy is not depression. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *10*(3), 314–319. https://doi.org/10.1176/jnp.10.3.314

Linehan, M. M. (1993). *Skills training manual for treating borderline personality disorder*. Guilford Press.

Louie, K., Glimcher, P. W., & Webb, R. (2015). Adaptive neural coding: From biological to behavioral decision-making. *Current Opinion in Behavioral Sciences*, *5*, 91–99.

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, USA*, *110*(15), 6139–6144.

Lynch, T. R., Rosenthal, M. Z., Kosson, D. S., Cheavens, J. S., Lejuez, C. W., & Blair, R. J. R. (2006). Heightened sensitivity to facial expressions of emotion in borderline personality disorder. *Emotion*, *6*(4), 647–655. https://doi.org/10.1037/1528-3542.6.4.647.

Mandelli, L., Petrelli, C., & Serretti, A. (2015). The role of specific early trauma in adult depression: A meta-analysis of published literature. Childhood trauma and adult depression. *European Psychiatry*, *30*(6), 665–680.

Marin, R. S. (1991). Apathy: A neuropsychiatric syndrome. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *3*(3), 243–254. https://doi.org/10.1176/jnp.3.3.243

Marin, R. S. (1996). Apathy: Concept, syndrome, neural mechanisms, and treatment. *Seminars in Clinical Neuropsychiatry*, *1*(4), 304–314.

Meyer, B., Johnson, S. L., & Winters, R. (2001). Responsiveness to threat and incentive in bipolar disorder: Relations of the BIS/BAS scales with symptoms. *Journal of Psychopathology and Behavioral Assessment*, *23*(3), 133–143.

Michely, J., Eldar, E., Martin, I. M., & Dolan, R. J. (2020). A mechanistic account of serotonin's impact on mood. *Nature Communications*, *11*, Article 2335. https://doi.org/10.1038/s41467-020-16090-2

Papalini, S., Beckers, T., & Vervliet, B. (2020). Dopamine: From prediction error to psychotherapy. *Translational Psychiatry*, *10*(1), 1–13.

Paulus, M. P., Feinstein, J. S., & Khalsa, S. S. (2019). An active inference approach to interoceptive psychopathology. *Annual Review of Clinical Psychology*, *15*, 97–122.

Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: Toward a synthesis and integrated model. *Annual Review of Clinical Psychology*, *10*, 393–423.

Rigoli, F. (2019). Reference effects on decision-making elicited by previous rewards. *Cognition*, *192*, Article 104034. https://doi.org/10.1016/j.cognition.2019.104034

Rigoli, F., Chew, B., Dayan, P., & Dolan, R. J. (2018). Learning contextual reward expectations for value adaptation. *Journal of Cognitive Neuroscience*, *30*(1), 50–69.

Rigoli, F., Friston, K. J., & Dolan, R. J. (2016). Neural processes mediating contextual influences on human choice behaviour. *Nature Communications*, *7*, Article 12416. https://doi.org/10.1038/ncomms12416

Rigoli, F., Friston, K. J., Martinelli, C., Selaković, M., Shergill, S. S., & Dolan, R. J. (2016). A Bayesian model of context-sensitive value attribution. *ELife*, *5*, Article e16127. https://doi.org/10.7554/eLife.16127

Rigoli, F., Rutledge, R. B., Dayan, P., & Dolan, R. J. (2016). The influence of contextual reward statistics on risk preference. *NeuroImage*, *128*, 74–84.

Robinson, T. E., & Berridge, K. C. (2001). Incentive-sensitization and addiction. *Addiction*, *96*(1), 103–114.

Robinson, T. E., & Berridge, K. C. (2008). The incentive sensitization theory of addiction: Some current issues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1507), 3137–3146.

Rosenthal, M. Z., Gratz, K. L., Kosson, D. S., Cheavens, J. S., Lejuez, C. W., & Lynch, T. R. (2008). Borderline personality disorder and emotional responding: A review of the research literature. *Clinical Psychology Review*, *28*(1), 75–91.

Rottenberg, J. (2005). Mood and emotion in major depression. *Current Directions in Psychological Science*, *14*(3), 167–170.

Rottenberg, J. (2017). Emotions in depression: What do we really know? *Annual Review of Clinical Psychology*, *13*, 241–263.

Rottenberg, J., Gross, J. J., & Gotlib, I. H. (2005). Emotion context insensitivity in major depressive disorder. *Journal of Abnormal Psychology*, *114*(4), 627–639. https://doi.org/10.1037/0021-843X.114.4.627

Rottenberg, J. E., & Johnson, S. L. (2007). *Emotion and psychopathology: Bridging affective and clinical science.* American Psychological Association.

Saudino, K. J. (2005). Behavioral genetics and child temperament. *Journal of Developmental and Behavioral Pediatrics*, *26*(3), 214–223. https://doi.org/10.1097/00004703-200506000-00010

Sauer, S., Walach, H., Schmidt, S., Hinterberger, T., Lynch, S., Büssing, A., & Kohls, N. (2013). Assessment of mindfulness: Review on state of the art. *Mindfulness*, *4*(1), 3–17.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *2*, Article 395. https://doi.org/10.3389/fpsyg.2011.00395

Smith, B. (2013). Depression and motivation. *Phenomenology and the Cognitive Sciences*, *12*(4), 615–635.

Starkstein, S. E., & Leentjens, A. F. (2008). The nosological position of apathy in clinical practice. *The Journal of Neurology, Neurosurgery & Psychiatry*, *79*(10), 1088–1092.

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, *106*(1), 5–15.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology*, *62*(6), 1041–1062.

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

Stewart, S. H. (1996). Alcohol abuse in individuals exposed to trauma: A critical review. *Psychological Bulletin*, *120*(1), 83–112. https://doi.org/10.1037/0033-2909.120.1.83

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*, *307*(5715), 1642–1645.

Volkow, N. D., Koob, G. F., & McLellan, A. T. (2016). Neurobiologic advances from the brain disease model of addiction. *New England Journal of Medicine*, *374*(4), 363–371.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience*, *5*(6), 483–494.

Woodford, M. (2012). Prospect theory as efficient perceptual distortion. *American Economic Review*, *102*(3), 41–46.