15

A Valuation Framework for Emotions Applied to Depression and Recurrence

Quentin J. M. Huys

Abstract

The burden of depression is substantially aggravated by relapses and recurrences, and these become more inevitable with every episode of depression. This chapter describes how computational psychiatry can provide a normative framework for emotions and an integrative approach to core cognitive components of depression and relapse. Central to this is the notion that emotions effectively imply a valuation; thus they are amenable to description and dissection by reinforcement-learning methods. It is argued that cognitive accounts of emotion can be viewed in terms of model-based valuation, and that automatic emotional responses relate to model-free valuation and the innate recruitment of fixed behavioral patterns. This model-based view captures phenomena such as helplessness, hopelessness, attributions, and stress sensitization. Considering it in more atomic algorithmic detail opens up the possibility of viewing rumination and emotion regulation in this same normative framework. The problem of treatment selection for relapse and recurrence prevention is outlined and suggestions made on how the computational framework of emotions might help improve this. The chapter closes with a brief overview of what we can hope to gain from computational psychiatry.

The Candidate Clinical Issue: Recurrence

One episode of an acute mental health illness is bad enough. Unfortunately, many psychiatric disorders are chronic, with episodes of wellness repeatedly punctuated by relapses in which symptoms reemerge. The relentless repetition is an important factor in the burden these disorders impose globally and individually, and it means that treatment of the early illness stages is crucial in avoiding an unremitting decline of well-being.

In this chapter, I focus on depression: a disease that is highly recurrent and imposes a heavy burden on the world's citizens (Whiteford et al. 2013). Of

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

those who experience a first episode of depression, around 40% will remain well and never go on to have a recurrence (defined as a reemergence of symptoms after a substantial period of absence) (Frank et al. 1991). The remaining 60% with a single episode will go on to develop further episodes. Thereafter, the proportion with unremittingly recurrent disease rises: 70% of those with two episodes will experience a third, and 90% of those with three episodes will experience a fourth (APA 2000). This process has been recognized since the time of Kraepelin. Hence, around 40% will have a single lifetime episode, at most 20% two episodes, 5–10% three episodes, and the remainder—a staggering 40% or more—will have more than three episodes (Angst 1992). At an average duration of 6–9 months per episode, this entails many ill years, and these numbers might even be optimistic (Hollon et al. 2006; Posternak et al. 2006). In addition, any one episode has about a 5–10% risk of becoming chronic, i.e., lasting for two years or more, with the risk for chronicity and recurrence being partially independent (Hollon et al. 2006). Hence, the early episodes are of fundamental importance as they mark a transition between those whose depression will have a relatively benign outcome and those who will experience a malignant course (Keller et al. 1983; Monroe and Harkness 2011).

Both antidepressant medication (ADM) and psychotherapy have proven utility in preventing relapses and recurrences. However, their differential indication is poorly understood, and it is likely that currently much can be gained from improving the targeting of treatments. For instance, DeRubeis et al. (2014) examined the ability to differentially predict treatment response to antidepressants and cognitive therapy based on standard clinical characteristics. Some variables were prognostic, predicting treatment response independent of modality (e.g., severity, age, IQ, chronicity), whereas others were prescriptive and differentiated between the modalities (e.g., marriage, employment, comorbid personality disorder, numbers of life stressors, and previous ADM trials). Using a simple general linear model combined with a standard machine-learning approach, DeRubeis et al. suggested that they could improve allocation so as to produce an average improvement of over 3 points on the Hamilton Rating Scale for Depression, which is clinically significant. McGrath et al. (2013) reported that insular metabolism measured by FDG-PET differentiated those who would respond to an ADM from those who would respond to psychotherapy. Others, however, have suggested that the subgenual anterior cingulate activity may differentiate those likely to respond to cognitive therapy from those likely to respond to ADMs (DeRubeis et al. 2008; Roiser et al. 2012). Using quantitative EEG to allocate patients to antidepressant treatment, DeBattista et al. (2011) reported an improvement of about 3 points on the Quick Inventory of Depressive Symptomatology—again, clinically a very significant finding. To date only a few attempts have been made to use such methods to address the problem of predicting relapse. Among these, the induction of dysfunctional attitudes with sad mood induction (Teasdale 1988; Segal et al. 1999, 2006) and variations in neuroimaging responses to sad movies (Farb et al. 2011) stand out. However, there appears to be no work on predicting specifically who will relapse after discontinuing ADMs. In addition to their clinical value, robust predictors of differential response might lead the way toward a novel understanding of the neurobiology of psychiatric disorders in terms of neurobiological features that are directly relevant for treatment.

Here, I propose that computational techniques are well placed to help us attain a better understanding of affective processes that are crucial in transforming depression from a mild and temporary into a chronic and disabling condition. I begin by introducing a computational framework for emotions that focuses particularly on model-based evaluations. Then I discuss aspects of model-based valuations relevant to depression relapse and attempt to outline a normative framework for phenomena such as stress sensitization and emotion regulation. Thereafter, a brief overview of the clinical management of relapses is provided, followed by specific candidate applications. I conclude with a critique of this approach and discuss its key limitations.

Before beginning, it is worth noting the distinction between recurrence and relapse: recurrences occur further apart in time and with more profound resolution of symptoms between the two exacerbations (Frank et al. 1991). The statistics of depressive episodes, however, do not reveal obvious points for such a categorization (Burcusa and Iacono 2007; Monroe and Harkness 2011). Instead, they point to degrees of risk for future exacerbations (Judd and Akiskal 2000; Judd et al. 2000; Dunlop et al. 2012). Variations in the definitions of relapse and recurrence are thus an important caveat, in terms of comparing and integrating across studies.

Valuation as a Framework for Emotions

Computational psychiatry provides broadly two approaches (Huys et al. 2016). The descriptive approach employs general, neuropsychologically agnostic methods to describe the relationship between neuropsychological measurements and some outcome of interest (e.g., applying regression to predict treatment response) (DeRubeis et al. 2014). The mechanistic approach is very different as it aims to describe how phenomena of interest come about using far more complex accounts of the data, ranging from biophysically realistic models of network dynamics to processes describing learning and decision making. Both of these complex accounts can be used to identify key variables, and by fitting models to data, these variables can be efficiently measured and improve the performance of descriptive approaches (Wiecki et al. 2015; Huys et al. 2016).

In approaching depression mechanistically, the first issue that needs addressing is how computational techniques could help explain such subjective and ethereal phenomena as emotions (Huys et al. 2011; Maia and Frank 2011; Montague et al. 2012). At the core of this suggestion is the link between

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

valuation and emotion (Huys et al. 2015a). Emotions are strongly characterized by subjective qualia which are easy to catalog but hard to comprehend. Following the animal literature, however, one can focus on their behavioral correlates and refer subjective aspects to future investigations, pending a better understanding of consciousness itself. The key argument is as follows: To the extent that emotions influence behavior, they implicitly implement a process that assigns some behaviors a higher value than others. This opens up the possibility to describe emotions by considering the processes that give rise to valuations. Accordingly, the subjective phenomena entering emotional awareness are introspective correlates of the primary valuation processes.

Accounts of valuation have benefited immensely from reinforcement-learning approaches (Sutton and Barto 1998), and nowhere most clearly than in the elucidation of the phasic dopamine signals in learning (Montague et al. 1996), the puzzling relationship of which was found to relate closely to a computationally precise learning signal. Reinforcement-learning techniques address one of two fundamental inferential problems that the brain faces: actions taken now influence which options will be available in the future. For instance, the house you buy now influences how easily you will be able to accept an exciting job offer at a faraway university next year. Hence, optimal behavioral choice requires the appropriate assessment of vast future consequences. Approaches to this problem come in at least two fundamentally different shapes (Sutton and Barto 1998) that are behaviorally, neurobiologically, and computationally discernible (Killcross and Coutureau 2003; Daw et al. 2005, 2011; Deserno et al. 2015). *Model-based valuation* depends on an understanding of the structure of the world. Choices are valued by inferring their future consequences according to one's understanding encapsulated in a model *M* that describes the consequences of taking specific actions in different situations. This requires processing power, but it is powerful and flexible. *Model-free valuation*, by contrast, assigns values to states or stimuli by virtue of their past association with rewards or losses. At the time of choice, model-free values are computationally cheap, but they demand substantial experience to be accurate. Hence, these two systems trade experiential for computational costs. One changes slowly with experience; the other does so rapidly but requires substantial cognitive resources.

Generally speaking, these two theoretical components can be mapped onto two broad currents of current thinking about emotions: automatic and cognitive accounts, respectively. The automatic account emphasizes that stimuli can activate emotional centers and dictate responses, largely foregoing any contact with cognition, such as when innately relevant stimuli result in reflexive approach, fight, flight, or fright responses. Hirsch and Bolles (1980), for instance, showed that innate defensive responses to predators can breed true over multiple generations without contact to the predator. In contrast, cognitive theories of emotions (e.g., Beck 1967; Lazarus 2006) argue that many emotions experienced by humans *follow* cognitive appraisals. The emotions

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

evoked by complex, ambiguous stimuli that are devoid of evolutionary importance depend on attributions and interpretations, linking them to causes that have meaningful consequences or relate to innately relevant stimuli. This insight arguably forms the cornerstone of cognitive therapies for a variety of psychiatric disorders (Beck 1967) and provides a way by which emotions (and their consequences) can be regulated.

Distinguishing between different types of values renders this argument more concrete. In Pavlovian scenarios, values $\mathcal{V}(s)$ are attached to stimuli, states, or situations s independently of behavior; in instrumental scenarios, $\mathcal{Q}(s, a)$ values are attached to a combination of stimuli (or states or situations) in combination with particular behaviors a; that is, values are attached to stimulus-action pairs s, a (Mowrer 1947; Dayan and Balleine 2002; Daw et al. 2005, 2006). In this theory, both of $\mathcal{Q}$ and $\mathcal{V}$ values can be derived through model-free or model-based mechanisms, leading to a quartet of values (Huys et al. 2014): model-based and model-free Pavlovian values $\mathcal{V}^{\mathrm{MB}}(s)$ and $\mathcal{V}^{\mathrm{MF}}(s)$, and model-based (MB) and model-free (MF) instrumental values $\mathcal{Q}^{\mathrm{MB}}(s, a)$ and $\mathcal{Q}^{\mathrm{MF}}(s, a)$. Importantly, however, only the latter $\mathcal{Q}$ values directly inform action choice. The former, Pavlovian $\mathcal{V}$ values, do not as they make no reference to actions a, only to stimuli s. The mapping from stimulus values s to actions depends on mappings $m^{\mathrm{Fix}}(s, a)$ between the stimuli and responses, which must be determined separately. Hence, while instrumental $\mathcal{Q}$ values can be modified to theoretically implement any kind of behavior, Pavlovian values $\mathcal{V}$ are effectively restricted to modulating otherwise specified fixed-response mappings $m^{\mathrm{Fix}}(s, a)$. Consider the innate link between positive valuation and approach. In a striking example of the possibly maladaptive nature of autoshaping, Hershberger (1986) had a hungry chick on a linear track facing a food tray. The food tray was mobile, and moved in the same direction as the chick, but at twice its speed. To get close to the food, the chick had to run away from the food as quickly as possible—an impossible task. This suggests that the stimulus food was innately linked to approach behavior; that is, $m^{\mathrm{Fix}}(\mathit{food}, \mathit{approach})$ had a high effective "value." Akin to Hershberger's chicks, humans are easily able to learn to emit a Go action for rewards and to withhold active responses to avoid losses, but they perform poorly when they have to choose No-Go to earn a reward or Go to avoid losses. The extent to which rewards interfere with No-Go, and losses with Go, is well captured by allowing the expected value of the stimulus, throughout learning to evoke Go and No-Go behaviors (Guitart-Masip et al. 2012).

Here I propose that model-based Pavlovian stimulus values might capture important aspects of the cognitive view of stimulus-bound emotional responses. Situations or stimuli are examined within a greater interpretative framework (the "model" $\mathcal{M}$). The resulting model-dependent valuation leads to the recruitment of particular species-specific "emotional" response patterns (and, as a correlate, also subjective qualities). Through the process of model-based valuation, a stimulus s might also become predictive of other future stimuli s'

and thereby come to recruit the fixed-response mappings associated with these. Second, conversely, model-free stimulus values $\mathcal{V}^{\mathrm{MF}}(s)$ might map better onto more automatic routes to acquiring emotional responses, whereby repeated exposure to a contingency between a conditioned stimulus and an unconditioned stimulus leads the conditioned stimulus to acquire the same value as the unconditioned stimulus, thereby triggering the same or similar innate response patterns by reference to past experience rather than by reference to an interpretative model. Third, it is important to consider the interaction between these processes. Finally, theories of valuation have been very useful in understanding the function of neuromodulators (i.e., the putative substrates of the majority of psychotropic medications). Due to space constraints, this will not be discussed here, and the reader is referred to existing work (Montague et al. 1996; Frank et al. 2004; Frank 2005; Yu and Dayan 2005; Niv et al. 2007; Dayan and Huys 2009; Cools et al. 2011).

Model-Based Valuation in Depression

This computational framework for emotions can capture facets of the cognitive features of depression, particularly those with relevance to relapse: hopelessness, stress sensitization, rumination, and emotion regulation.

Stress and Hopelessness

Stressors have an important and causal role in the onset of depressive episodes (Kendler et al. 1999, 2000). Just like initial episodes, recurrences can also be provoked by stress: subjects who score higher on measures of severe life events after recovery have a higher risk of enduring a recurrence (Monroe et al. 1986). However, the experience of depression also leaves a "scar" (Burcusa and Iacono 2007) and stress maligns more with every further episode. Specifically, while the first episode is strongly associated with prior severe life events, subsequent episodes are less dependent on a severe life event (Kendler et al. 2001) and occur more autonomously or with less severe events (Kendler et al. 1995; Wichers et al. 2007). In addition to this "kindling effect," depression also influences people's behavior such that they self-select into high-risk situations or manage difficult situations poorly. In an effect known as *stress generation* (Hammen 1991), with every episode the rate of further severe life events increases (Harkness et al. 1999; Kendler et al. 2001; Liu and Alloy 2010); this increases the risk of further episodes, even in the absence of sensitization.

Stress arises from an interaction between stimuli that are potentially threatening and individual situational appraisals (Lazarus 2006). Whether any one particular stressor (e.g., a relationship breakup) results in a depressive episodes depends on the individual's perception of it (it might come as a relief!). Thus models which formalize the complex impact of stressors play an important role in our understanding of depression, and several theories of depression have

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

emphasized that depression is characterized by cognitive features (Abramson et al. 1989; Alloy et al. 1999) that have the potential to transform stressors malignantly. A particularly useful model is that of learned helplessness (Maier and Watkins 2005), where healthy animals are exposed to controllable or uncontrollable stress. Depression-like signs and symptoms only arise if the shock is perceived as uncontrollable, and this is learned from the experience of the shocks. In terms of the current argument, these inferential processes can be readily captured through the use of model-based valuation $Q^{\mathrm{MB}}(s, a)$ (Huys and Dayan 2009; Lieder et al. 2013). As animals experience the shocks, they also learn about the relationship between shocks and behavior. To the extent to which this learning involves modifying a general prior belief about how controllable desirable outcomes are, this experience will generalize to model-based valuations in other situations. Normatively speaking, generalization between different situations should depend on similarities between these situations. One similarity is the presence of the agent, and hence inferences about the agent's own abilities support generalization and are naturally global, stable, and internal, which are the core features of hopelessness (Abramson et al. 1989).

The influence of a prior on controllability in model-based valuation also provides reasonable accounts for depressive symptoms more generally. A belief about a lack of control impacts the expected exploitability of options in the world, and thereby the opportunity cost for behavior. This variable, which has been linked to tonic dopamine (Niv et al. 2007; Hamid et al. 2016) and affects the level of energy expenditure, is a reasonable substrate for the core components of perceived loss of energy, psychomotor retardation, and diminished drive, all of which have similar discriminatory power as anhedonia (McGlinchey et al. 2006; Mitchell et al. 2009). A wealth of other features of reward and loss sensitivity also attests to the involvement of model-based valuation and has been reviewed elsewhere (Huys et al. 2015a).

The modification of beliefs and their influence on valuation in model-based systems provides a formalization of cognitive theories of depression. In terms of relapse, this formulation could also account for the "scar" left by previous episodes. As the prior is updated and learned, it becomes stronger and can now be combined with a weaker likelihood term (fewer data) to provide strong beliefs that a particular situation is uncontrollable. Such a process may capture the finding that the dependence of depressive episodes on severe life events weakens because depressive episodes would more easily be triggered by weaker events (Kendler et al. 2001). A helpless prior belief would also impair the ability to exploit opportunities in the environment, both in terms of reward seeking and loss avoidance. If the immediate future is poorly controllable, the more distant future is entirely uncertain and the immediate cost prevails over long-term gains (Huys et al. 2009). This impairs the ability to make appropriately farsighted decisions and would particularly hurt in difficult situations that demand careful, farsighted behavior, hence increasing the risk of further severe life events (Hammen 1991; Liu and Alloy 2010).

This account hinges on the impact of stress on recurrence being mediated by the worsening of cognitive biases, which is a controversial topic. On one hand, cognitive distortions measured by the Dysfunctional Attitudes Questionnaire appear to be sensitive to the depressive state, to recover after an episode (Teasdale 1988), to have predictive value for recurrence after induction of sad mood (Segal et al. 1999, 2006), and to interact with stress in predicting depressive onset (Lewinsohn et al. 2001). On the other, negative cognitive styles—a more direct measure of helpless thoughts and beliefs (Alloy et al. 2000; Haeffel et al. 2008)—mediate medium-term effects of stressful experiences on depressive symptoms in (Haeffel et al. 2007), differ between remitted and never depressed subjects (Haeffel et al. 2005), and predict longitudinally depressive symptoms (Iacoviello et al. 2010; Pearson et al. 2015). In the study by Lewinsohn et al. (2001), however, attributional style predicted depression only for mild levels of stress, and neither this nor dysfunctional attitudes interacted with a history of depression in predicting onset. Hence, cognitive styles motivated by helplessness theory appear, overall, to satisfy these requirements, although whether and to what extent they mediate the kindling effects of stress or the self-generation of stress is as yet unknown.

Metareasoning, Rumination, and Emotion Regulation

The computational models discussed thus far have made little reference to the specifics of the algorithmic implementation of the biases, but have just described their overall existence and consequences. One overarching point is that model-based inference is computationally extremely costly. Because the future ramifications of current choices are so exponentially vast (looking d steps ahead when there are n options at every steps involves evaluating n^d combinations), their extensive consideration either requires (a) unreasonably large computational resources, thus relying on a plethora of approximations and shortcuts (Huys et al. 2012, 2015c) that are potentially relevant for psychiatry (Huys et al. 2015b), or (b) sacrifices to be made in terms of some of the key features of model-based cognition (Daw and Dayan 2014).

In addition, this resource constraint leads to the metareasoning problem of how to allocate internal computational resources optimally (Anderson and Oates 2007; Hay and Russell 2011). The subject faces both an external problem of to evaluate actions, and an internal problem of how to allocate cognitive resources to the evaluation of particular behavioral options (e.g., to find the best action). The latter problem implies a decision problem about the decision problem: having considered a particular option, what option should next be considered?

This view of metareasoning allows us to consider further aspects of depressive cognition, including rumination, emotion regulation, and cognitive control, in the formal framework of valuation. Rumination is an internal predilection to focus on the causes, meanings, and consequences of depressive symptoms (Nolen-Hoeksema 1991) and is driven in part by the kinds of features which drive helplessness, being particularly prominent in those with high level of chronic burden (e.g., housework, caring for children or elderly) and a low sense of mastery (Nolen-Hoeksema et al. 1999). Hence, this raises the tantalizing possibility that a prior belief about lack of control could render all possible options equally unappealing (Huys and Dayan 2009), and thereby impair the ability to identify useful options to consider, with attempts to identify a valuable option resulting in a persistent preoccupation with the status quo.

Aspects of emotion regulation can be viewed similarly. Emotion regulation refers to cognitive, behavioral, and other strategies used to control which emotions are experienced (Gross 1998). Reinterpretation, for instance, turns a half empty glass into a half full glass. If emotions are viewed as the (automatic) recruitment of innate fixed-response patterns $m^{\mathrm{Fix}}(s, a)$ by model-based valuations $\mathcal{V}^{\mathrm{MB}}(s)$, then emotion regulation can either refer to the modulation of the link between valuation and the response pattern, or to the modulation of the valuation itself. The former would involve altering the metareasoning selection process, for instance by focusing internal evaluations more on those values likely to be positive. This distinction would parallel that drawn between response-focused and antecedent-focused emotion regulation: Gross (1998) has argued that the former is maladaptive in terms of a wide variety of emotional correlates, whereas the latter is adaptive. Viewing them as an internal focus on different valuations provides one rationale for why this might be. The focus on positive values increases the chances of selecting behavioral patterns with positive consequences, while the latter does not and, if anything, possibly impairs evolutionary set automatic adaptive responses (see also Coan and Allen 2007).

Interestingly, some of these metareasoning strategies seem to be both directly accessible through conscious report (Papageorgiou and Wells 2002, 2003) and amenable to treatment (Wells et al. 2012).

Algorithmic Accounts of Model-Based Evaluations: Pruning and Memoization

The precise processes which underlie model-based evaluation are currently of great interest in neuroscience (Johnson and Redish 2007; Pfeiffer and Foster 2013; Daw and Dayan 2014; Doll et al. 2015; Kurth-Nelson et al. 2015). By formalizing processes such as rumination and emotion regulation in the valuation framework, one can hope to benefit from these advances. I note, however, that these investigations are in their infancy. Nevertheless, because the internal evaluation processes are so rich and complex, computational modeling of decision making might come to be a helpful tool.

Detailed computational models of planning have identified at least two aspects in the regulation of the internal search that are potentially relevant for depression: pruning and memoization. When subjects are asked to plan ahead, they partially solve the metareasoning problem by reflexively pruning; that is, they stop thinking about an option when that option raises the possibility of a salient loss, even when forfeiting larger rewards behind the large losses (Huys et al. 2012). This internal thought inhibition has been suggested to be analogous to behavioral inhibition and related to serotonin (Dayan and Huys 2008; Crockett et al. 2012; Geurts et al. 2013). Preliminary results suggest that it involves the subgenual anterior cingulate cortex, a substrate known to be important in rumination (Sheline et al. 2009) as well as treatment response (Mayberg 2009; Fu et al. 2013), and to mediate the impact of aversive events on choices (Amemori and Graybiel 2012). Pruning might also be related to the notion of inhibitory control in depression, which refers to the ability to inhibit the processing of aversive information (Gotlib and Joormann 2010). Healthy controls show an impairment or delay in processing an affectively negative target that was a distractor on the previous trial, but this is absent in depressed patients (Joormann 2004, 2006).

Memoization refers to the reuse of results from previous cognitive efforts. Rather than recomputing the solution to a problem previously faced, individuals tend to reuse the previous solutions (Huys et al. 2015c). This provides a way for the results of cognitive processes to ingrain themselves very rapidly.

Both pruning and memoization have yet to be shown to be directly implicated in depression. Nevertheless, they provide potential algorithms for the implementation of theories of depression that rely on the interaction between two types of processes: one high-level cognitive factor and another more low-level attentive or more reactive factor. For instance, the two-factor model of relapse (Farb et al. 2015) and the cognitive neuropsychological model of depression (Roiser et al. 2012) both suggest that depression involves the integration of two components: (a) negative affective biases are present in terms of attention, memory, and perception, and are risk factors for the development of a first episode; (b) changes in these affective biases precede treatment response (e.g., Wells et al. 2014). The theory suggests that these biases are, over time, consolidated into negative schemata and rumination, but it does not explain how this process might occur. The impact of Pavlovian reflexes on metareasoning is one path by which affective biases might influence cognitive processes, and memoization might provide a rapid way for stamping in the results.

Valuation might thus provide a normative framework for emotions, and investigations into model-based computation and metareasoning might help to integrate cognitive components of depression. This may potentially account for stress sensitization, helplessness, rumination, and emotion regulation.

Clinical Interventions for Recurrence

As stated, both psychopharmacological and psychotherapeutic interventions have proven useful in preventing relapses and recurrences of depression. Here

I briefly review features of ADMs and psychotherapy which aid in the management of recurrent depression.

Antidepressant Medications

ADMs have shown convincing efficacy in reducing the risk of relapse or recurrence of depression. Several meta-analyses (Viguera et al. 1998; Geddes et al. 2003; Kaymaz et al. 2008; Glue et al. 2010) estimate the reduction in the odds of a relapse due to continuation of antidepressants at around 70%, with a relative risk reduction at around 50% and an absolute risk reduction at 20% (from 40%); the number needed to treat (NNT) to avoid one additional relapse is thus 5 over one year (Geddes et al. 2003). These findings have led to recommendations of relatively long treatment periods (e.g., Anderson and Pilling 2010; Bauer et al. 2013), informed by variables describing the course of the illness, mainly severity, the number of prior episodes, and the duration of treatment (Härter et al. 2009; Anderson and Pilling 2010; Bauer et al. 2013). As second and third generation ADMs are tolerated well (Anderson and Tomenson 1995; Cipriani et al. 2009), chronic treatment is a feasible possibility.

Though none of these outweigh the strong current evidence in favor of treatment, it is important to remember that there are only very few trials that have looked at medication on a timeframe of more than two years, and that the impact of the number of past episodes may hide a variety of other variables (e.g., lifetime stress exposure, ADM treatment history, and residual symptoms; see Appendix). The fact that relapse rates after discontinuation of ADMs show a very early excess that slowly drops off (Viguera et al. 1998; El-Mallakh and Briscoe 2012; see Appendix) has fuelled arguments that the antidepressant effect might either wear out (Rothschild et al. 2009) or result in long-term withdrawal syndrome (Fava et al. 2015). There are also reports that long-term treatment could lead to the development of adaptations that render future episodes more likely, particularly after discontinuation. For instance, there are meta-analytic reports that the relapse rate is higher among patients who have responded to antidepressants, than those who have had a spontaneous remission. In fact, the stronger the *in vitro* perturbational effect of the antidepressant, the stronger the effect, even when controlling for the number of prior episodes, treatment duration, and the stringency of recovery definition (Andrews et al. 2011). Hence, it cannot be excluded that adaptations to the antidepressants themselves contribute to the excess relapses seen very early after discontinuation (see Appendix). Such phenomena raise questions about the practice of chronic treatment. Furthermore, antidepressants are also not devoid of other adverse side effects, which range from impairments of sexual function to increased risks of hemorrhage and cardiac arrhythmias. They are discontinued at extremely high rates, with a half-life as low as 20 days; up to 75% of patients discontinue ADMs within the shortest recommended treatment duration (after a first episode) of six months (Olfson et al. 2006; Lee and Lee 2011). Finally, an NNT of 5 is clinically attractive, but chronic long-term medication does place a nonnegligible burden on four out of the five patients who will not benefit from the medication.

Hence, improving the targeting of ADMs to those who are most likely to benefit from continuation or prophylactic treatment would be extremely valuable. Those who will not benefit can be offered alternative preventions, or, if they are at low risk, followed up without intervention. Better indications for treatment might increase the concordance with treatment among those prescribed an ADM, and reduce the number of those treated in vain.

ADMs prominently impact both serotonergic and noradrenergic neuromodulation, and both have been the target of computational models of the type described here. Serotonin has been suggested to relate to automatic behavioral inhibition (Soubrie 1986; Deakin and Graeff 1991; Dayan and Huys 2009; Crockett et al. 2012). However, recent advances in optogenetics have raised doubts about such accounts (Cohen et al. 2015; Dayan and Huys 2015).

Psychotherapy

Psychotherapy is approximately as effective as ADMs in the acute treatment of depression, resulting in remission in 50–60% of patients (DeRubeis et al. 1999). The relapse rates after treatment (approximately 50% over two years; Vittengl et al. 2007) are also high. When compared directly to treatment with ADM, however, psychotherapy appears to have a longer-lasting effect: relapse rates after cognitive therapy, mindfulness-based cognitive therapy (MBCT), or behavioral activation therapy are similar or non-inferior to those under continued ADM treatment (Blackburn et al. 1986; Evans et al. 1992; Teasdale et al. 2002; Fava et al. 2004; Hollon et al. 2005).

Continuation cognitive therapy after remission under either cognitive therapy or antidepressants also reduces relapse rates compared to switchover to placebo (Jarrett et al. 2001; Segal et al. 2010) and effects are similar to those of continuation ADM therapy: approximately 20% absolute risk reduction (Geddes et al. 2003; Piet and Hougaard 2011). Unlike acute-phase psychotherapy, continuation psychotherapy does not seem to have a longer-lasting effect after termination (Jarrett et al. 2013). Adding cognitive therapy to flexible pharmacotherapy may help to achieve recovery, defined as a longer-lasting remission (Hollon et al. 2014). Whether relapse rates after recovery with psychotherapy are higher than after spontaneous recovery, as is true for antidepressants (Andrews et al. 2011), does not appear to have been examined.

As with ADMs (Viguera et al. 1998), there are indications that patients who experience multiple past episodes benefit more from psychotherapy. MBCT, for instance, has a stronger impact on patients with ≥ 3 episodes than those who experience at most 2 episodes. It also appears to be more efficient for treating recurrences that are not initiated by life events (Ma and Teasdale 2004). In fact, MBCT was specifically designed with relapses in mind (Teasdale et al. 2002).

Computational Psychiatry Approaches to Prevent Recurrence

Both psychotherapy and ADMs have well-proven utility in the prevention of relapse and recurrence. However, both appear to have to be administered chronically and both have NNTs around 5, meaning that a substantial portion of those given chronic treatment have little to no benefit. Furthermore, there are drawbacks to both therapies, and individuals discontinue, in particular, ADM therapy at a very high rate. Better allocation to different interventions should improve outcomes and reduce the development and establishment of recurrent depression.

Descriptive Atheoretical Approaches

A first important task for computational psychiatrists is to amalgamate the extant information and use it to derive individual risk scores. What is the relapse risk for a patient who wishes to discontinue medication six months after his second episode, who responded to the first medication, and who has had only some sleeping problems for the past few weeks? The aim in this would not be to necessarily identify particular risk factors, but rather to use agnostic statistical tools to amalgamate a variety of variables into a robust prediction. The literature provides very rich hints as to which variables are most likely to prove important and informative. However, what is as yet not well understood is which variables provide incremental predictive power in which combinations. By identifying individual variables that modulate relapse risk in individual patients, such predictions might be able to guide treatment (which, of course, would have to be assessed specifically). Multiple predictors exist for response to a particular treatment mode, ranging from clinical descriptors (DeRubeis et al. 2014) and EEG measures (Pizzagalli et al. 2001; Mulert et al. 2007; Korb et al. 2009; Leuchter et al. 2009), emotional blunting (Peeters et al. 2010), subgenual resting state activity and response to aversive stimulation (DeRubeis et al. 2008; Roiser et al. 2012; Siegle et al. 2012; Fu et al. 2013) and others. These techniques have not yet entered clinical practice, possibly because they do not really address the actual clinical problem (i.e., treatment allocation). In the setting of relapse prevention, they may prove useful to assess continuance of a specific treatment. However as yet, it is entirely unclear whether the same measures that predict initial treatment response have any value in predicting maintained response.

To begin, purely descriptive, atheoretical machine-learning approaches should be applied to the existing features, to combine them and produce better predictors of an individual's relapse risk (DeRubeis et al. 2014). A next step would be to apply this approach to the differential prediction of relapse under pharmacological versus psychotherapeutic management, and then to the differential relapse risk with different specific treatments. To the extent that such data exists, one could even imagine extending this approach to other interventions which psychiatrists deploy, such as supported living, employment, or psychoeducational interventions.

Mechanistic Approaches

It is an open question to what extent mechanistic approaches will be helpful, in particular, in the valuation account of emotion. Nevertheless, two areas stand out as potential applications. The first is the differential indication for psychotherapy. Psychotherapies are built around specific conceptualizations. For instance, MBCT has extracted a particular component of cognitive therapy, which is the mindful internal distancing from emotions. There is good evidence that the therapy results in changing mindfulness and that this mediates improvement (Teasdale et al. 2002; Ma and Teasdale 2004; Kuyken et al. 2010), but whether a dysfunction (low mindfulness) is a differential indicator is, to my knowledge, not yet settled. In fact, to the best of my knowledge there are no validated measures that predict differential response to different types of psychotherapy. As I argue here, computational notions of valuation provide an integrative framework for emotions and might accommodate several processes that play salient roles in psychotherapy. It might thus be fruitful to build detailed computational models of the active ingredients of psychotherapy. To the extent to which this turns out to be feasible, it might also allow targeted measurements to support differential treatment allocation.

The second area concerns the examination of how dynamic interactions between valuation factors establish and stabilize high-risk states for relapse (i.e., how a single episode is being transformed into a chronic disabling condition). This includes (a) the interaction between helpless cognitions and guided exploration and exploitation, (b) the impact of aversive Pavlovian thought inhibition on mood regulation and on the ability to recognize worthwhile longer-term effort investments, and (c) the impact of memoization on the establishment of automatic thought patterns and more. Building models to explain how these processes result in the establishment of recurrent or chronic disease states might help us identify key variables as well as design efficient measures to manage depression.

Discussion

Psychiatry is a field of medicine that encompasses the extremely complex phenomena of subjective suffering in society and the high-level functions of the human brain. Suffering and social function are hard outcomes to measure reliably or in an appropriately objective manner, and this is arguably at the heart of difficulties in providing hard definitions of illness in psychiatry. Nevertheless, the aim of computational approaches to psychiatry must ultimately be to improve patient outcomes, and hence to improve subjective suffering or functioning in society. Taking a step back, this is most likely to be achieved if the approach addresses specific clinical issues and measures quantities that are directly relevant to care decisions. This could arise directly through better targeting of existing interventions or the development of novel ones. Alternatively, computational approaches might lead to improved diagnoses, in the sense that the diagnoses would encapsulate an improved understanding of the problem at hand and thus lead to a better allocation or development of interventions.

Broadly speaking, mental health practitioners avail themselves of at least five types of interventions:

1. Social support, which defines and validates illness for various purposes (e.g., sick leave, defense arguments in legal issues), provides housing and financial as well as provision and maintenance of support networks
2. Educational efforts, which informs patients, their caregivers, and support networks about the illness and its potential consequences as well as manages stigma and aids understanding of available treatment options; other aspects (e.g., emotional regulation skills, social skills training) may also be relevant
3. Psychotherapy
4. Psychopharmacology
5. Interventions such as electroconvulsive therapy or surgery, which are primarily aimed at patients whose symptoms do not respond to other treatment options

As they seek to improve the conditions of patients, public health workers could benefit from more effective means of assessment and decision making Although the computational tools reviewed here may not directly improve social support measures or psychoeducational approaches, or yield novel molecular targets for medications, they could be crucial in targeting these resources.

Computational psychiatry encompasses two broadly distinguishable approaches. One approach is through mechanistic techniques, which aim to understand the system at hand. Limits to this approach include a dependence on costly data acquisition and difficult learning tasks. Although such tasks can powerfully probe internal processes and mechanisms of cognition, they are often not widely deployed in the clinic because of their temporal costs; they require very substantial attentional resources and cooperation and are poorly adapted to the severely ill patient. They also require substantial expertise to conduct. The other approach involves purely descriptive discovery. This involves agnostic statistical or machine-learning techniques to discover patterns or structure in informative data sets. In terms of the three ways computational techniques could be clinically helpful, it appears that diagnostic refinement and possibly the development of novel interventions might rely more on mechanistic advances, whereas treatment allocation (particularly among treatments with similar rationale) might benefit more imminently from an atheoretical approach.

There are many glaring omissions in this chapter (e.g., the absence of a discussion on neuromodulators). Reinforcement-learning models have provided deep insights into the function of neuromodulators (Dayan 2012) and might provide useful measures on the state of neuromodulatory systems, which could help in the differential allocation of treatment. This is, however, a very tall order, as models are not refined enough at present to be able to distinguish between subtle differences in current pharmacological treatments (though see Maia and Cano-Colino 2015). Here, descriptive methods might guide future advancement (DeBattista et al. 2011).

Conclusion

In this chapter, I have set forth a formal framework for emotions and described how some aspects of dysfunction can be conceptualized using it as well as how it might provide mechanisms for inferring internal variables that determine emotional experience. Descriptive methods are most likely to provide tools to target treatments more efficiently in the near future. The hope is that mechanistic models might yield results by capturing specific processes in a more focused manner, thereby allowing features that are critical to treatment and disease progression to be measured efficiently.

Acknowledgments

I would like to thank the organizers for the opportunity to write this paper. I gratefully acknowledge extensive discussions with Isabel Berwian, Peter Dayan and Nathaniel Daw on these and related issues. This work was partly funded by a Swiss National Science Foundation grant (SNF 320030L_153449).

Appendix

This supplement presents further details on relapse after antidepressant medication discontinuation. The number of previous episodes has a very profound impact. Figure 15.1a, b show the survival curves after medication discontinuation and continuation, respectively, as a function of the number of prior episodes from an early meta-analysis (Viguera et al. 1998), echoed by more recent ones (Kaymaz et al. 2008). However, the number of previous episodes hides substantial individual variability.

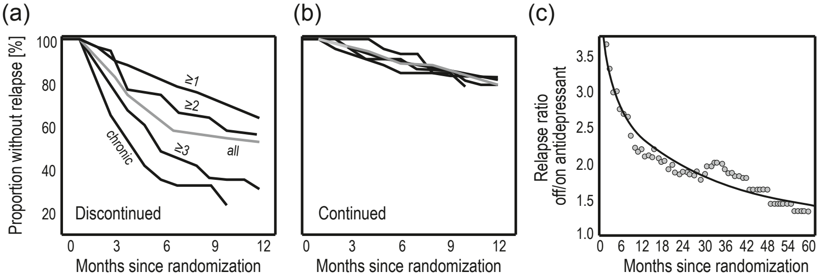

Figure 15.1 Relapse after antidepressant discontinuation. (a) Survival curves over one year after after randomization to placebo as a function of the number of previous episodes. Patients with more previous episodes relapse at a higher rate. (b) Survival curves over one year after continuation of antidepressant as a function of previous episodes. Relapse rate no longer depends on previous history. (c) Five year ratio of relapse risk as a function time since antidepressant discontinuation, compared to relapse risk when continuing medication. The risk is relatively high early on, and over a period of several years slowly falls toward one. Note though that the right tail of the function relies on few studies with small sample sizes. Figure adapted from Viguera et al. (1998).

First is the stage of the depressive disorder at the time of randomization. The probability of treatment response worsens with chronicity (Rush et al. 2006a), symptom severity, and difficult treatment (e.g., the number of treatment steps required to achieve remission) (Kennedy et al. 2003; McGrath et al. 2006; Rush et al. 2006a; Burcusa and Iacono 2007; Patten et al. 2012). Hence these results likely rely on a progressively more selected subset of patients who have responded despite increasingly recurrent disease. Importantly, residual subthreshold symptoms themselves predict early relapses, both in placebo-controlled trials (Andrews et al. 2011; Dunlop et al. 2012) and naturalistically, even after the first episode (Judd et al. 2000); they also can be a stronger risk factor than the number of past episodes (Judd et al. 1998). As there is no accepted definition of remission or recovery (Monroe and Harkness 2011), meta-analyses struggle to control fully for the extent of residual symptoms. Longer and deeper recovery predicts lower recurrence rates, but one important question is whether those with better or worse recovery benefit more from continued treatment (Segal et al. 2010; Dunlop et al. 2012). Importantly, the length of treatment itself does not appear to affect relapse rate, remaining essentially unaltered over courses of medication from one to two months to one year (Geddes et al. 2003; Kaymaz et al. 2008; Glue et al. 2010; Andrews et al. 2011).

However, patients with longer histories of depression typically have had more extensive, longer exposure to antidepressants, possibly leading to long-term alterations which might increase risk after antidepressant discontinuation. Figure 15.1c shows that the relative risk for a relapse occurs very early on and slowly approaches unity over five years. The temporal proximity of the excess risk to discontinuation has raised questions (El-Mallakh and Briscoe 2012) as to its origin and possible confounds, such as partial unblinding (with subjects noticing that the drug might have been been discontinued), a “wearing off” or tachyphylaxis of antidepressant function (Rothschild et al. 2009), or delayed withdrawal effects though drug half-life and discontinuation mode have somewhat unreliable effects on relapse rates (Viguera et al. 1998; Kaymaz et al. 2008; Andrews et al. 2011; Fava et al. 2015). For instance, in the PREVENT trial, which compared venlafaxine to fluoxetine for relapse prevention, continued treatment appeared to have an increased beneficial impact after two years compared to one year (Keller et al. 2007; Kocsis et al. 2007). Such findings speak to the notion of oppositional tolerance, whereby antidepressant discontinuation leads to a relapse rate that is heightened above the “natural” rate. Andrews et al. (2011) found that the relapse rate, after achieving remission on placebo, is lower than after achieving remission on antidepressant and discontinuing it. Importantly, these results hold when controlling for depression history and for recovery definitions. They also found that the extent to which the antidepressant discontinuation relapse rate is increased is proportional to how strongly each medication influenced neuromodulatory systems in animal models, arguing that adaptive processes may increase the risk of relapse independently of any depression disease processes.

From “Computational Psychiatry: New Perspectives on Mental Illness,”
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.