Better safe than sorry? - An Active Inference Approach to Biased Social Inference in Depression

Lukas Kirchner[1*], Anna-Lena Eckert[2*], Max Berg[1], Dominik Endres[2], Benjamin Straube[3], Winfried Rief[1]

[1]Department of Psychology, Clinical Psychology and Psychotherapy, Philipps-University of Marburg

[2]Department of Psychology, Theoretical Cognitive Science, Philipps-University of Marburg

[3]Department of Psychiatry and Psychotherapy, Philipps-University of Marburg

Authors Note

[*]Both authors contributed equally to this work.

Correspondence concerning this article can be addressed to: Lukas Kirchner, Department of Psychology, Clinical Psychology and Psychotherapy, Philipps-University of Marburg, Gutenbergstraße 18, 35037 Marburg, Germany, Email: lukas.kirchner@uni-marburg.de, Phone: +49 6421-2828904.

Word count: 5044

Note: This manuscript is currently in submission and has not been approved for publication yet.

1. Depression: a social disorder

Since Coyne's (1976) influential characterization of depression as an interactional disorder, there has been a large body of research on the social factors involved in the development and maintenance of depressive symptoms. By now, there is broad consensus that individuals with depression show marked distortions in the anticipation, perception, and processing of social information, as well as in their social behavior. For example, depressed individuals are more sensitive to rejection (Gao et al., 2017) and perform considerably worse at tasks measuring theory of mind (Nestor et al., 2022); they also reveal poor social problem-solving when interacting with others (e.g., Hames et al., 2013).

While social impairments are well described in depression (for reviews, see Bird et al., 2018; Hames et al., 2013; Kupferberg et al., 2016; Segrin, 1990, 2000; Starr & Davila, 2008; Tse & Bond, 2004; Weightman et al., 2014), a computational account is lacking of how these impairments differ from healthy social functioning and why they develop, are maintained, and manifest in such different ways (Smith, Badcock, et al., 2021). Because of its high degree of formalization and flexibility, a computational model of these processes would enable researchers to functionally link heterogeneous social symptoms in depression to various biases in patients' internal social belief systems, which would have important implications for diagnostics, etiology, and tailored interventions (Kirchner, Eckert, et al., 2022; Schwartenbeck et al., 2015; Smith, Badcock, et al., 2021).[1] In this paper, we describe the basic architecture of the model we have in mind – and focus on its motivation and functional form (as opposed to presenting numerical analyses with a particular example). Possible application and operationalization ideas will be described in the "Implications and Future Research" section.

From an etiological perspective, research from developmental and clinical psychology suggests a bidirectional vicious cycle between depressive symptoms and impaired social functioning (e.g., Beeson et al., 2020; Kirchner et al., 2022), with social trauma, such as peer rejection, childhood maltreatment, and lack of social support as common antecedents and risk factors (Gariépy et al., 2016; Nanni et al., 2012; Platt et al., 2013). In this view, social trauma and the lack of social support can lead to socially maladapted perception (e.g., rejection sensitivity, Zimmer-Gembeck, 2016), informational processing (e.g., interpretation biases, Everaert et al., 2017) and behavior (e.g., maladaptive interpersonal styles, Bird et al., 2018; Evraire & Dozois, 2011; Mikulincer & Shaver, 2012), which leads to the generation of interpersonal stress (e.g., rejection, Starr & Davila, 2008) that eventually promotes depressive symptoms by impairing social imperatives (e.g., sense of belonging and social connectedness, Bishop et al., 2022; Cruwys et al., 2014). Depressive symptoms, in turn, can lead to worsening social skills (e.g., as a result of social withdrawal, Trew, 2011), which can in turn result in negative social interactions and new social trauma, or in losing social support (e.g., Dobson et al., 2014).[2]

Concerning the functionality of the (social) symptoms of depression, evolutionary psychologists suggest that depressive symptoms may represent entirely adaptive responses to exceptionally aversive social contexts (at least in the short term), which minimize uncertainty in the social environment (e.g., "better-safe-than-sorry", Badcock et al., 2017). Adaptive processes like the mobilization of social support may be triggered (Watson & Andrews, 2002). However, when applied in the long term and regardless of changes in the social context, these strategies become maladaptive (Rantala et al., 2018), with serious consequences for those affected (e.g., suicidality, Stellrecht et al., 2006).

[2] Note that this aetiological model is particularly applicable to chronic forms of depression, where interpersonal trauma and interpersonal problems are core characteristics (Schramm et al., 2020).

2. On the maintenance of social symptoms in depression

One important question is why social perception and behavior do not normalize when the social context improves. Explanations from general and clinical psychology suggest several mechanisms are involved in this phenomenon - for example, learned helplessness (Seligman, 1972), neurobiological dysfunction (Porcelli et al., 2019), greater neural effort processing social cues (Suffel et al., 2020), pre-operational thinking (J. P. McCullough, 2003), dysfunctional core beliefs (Beck, 1963, 1964), negative relational schemas (Baldwin, 1992), increased elaboration of negative information (Kasch et al., 2002), blunted elaboration of positive information (Chen et al., 2015), persistent negative expectations (Kube et al., 2020; Roepke & Seligman, 2016), and cognitive immunization (Rief & Joormann, 2019).

Although we believe all these factors play an important role, we suggest that it is primarily *how* (and *by what model* of the social world) individuals with depression infer the characteristics of social contexts *through social action* (e.g., social decision making) that contributes to the maintenance of these symptoms (compare with “assimilation”, Panitz et al., 2021). Thus, clinical observations indicate that individuals with depressive symptoms (and specifically those with a social-trauma history) reveal distortions in exploring their interpersonal context before making social decisions (e.g., choosing to withdraw from a social situation). Instead, they stick to schema-congruent social strategies (e.g., "better-safe-than-sorry", Badcock et al., 2017) that prevent new learning and promote “fitting” schema-congruent experiences ("As I said, no one talked to me at the party"). This notion is supported by research linking depression to social withdrawal (e.g., Seidel et al., 2010), highly inhibited behavior (e.g., Kasch et al., 2002), low social information seeking, passivity and social avoidance (e.g., D’Iuso et al., 2018), as well as higher levels of distrust, shame, and disappointment in social situations (e.g., Fernández-Theoduloz et al., 2019). There is also evidence that depressed people have difficulty drawing the right conclusions from social interactions (for an overview, see

Weightman et al., 2014), and demonstrate aberrant patterns of reward learning and decision making in social contexts (Safra et al., n.d., 2019).

Clinicians from different theoretical backgrounds are well aware of this problem, and try to improve the inference of social contexts through psychoeducation (e.g., Cuijpers et al., 2009), interpersonal context discrimination training (e.g., McCullough et al., 2014; McCullough, 2006a), behavioral activation (e.g., Uphoff et al., 2020), experiential and emotion-focused techniques (e.g., Renner et al., 2013; Roediger et al., 2018), role disputes (e.g., Klerman & Weissman, 1994), training new interpersonal skills (e.g., Bellack et al., 1996; Lipsitz & Markowitz, 2013), repeated social exposure (e.g., Muñoz et al., 2000), disciplined personal involvement (e.g., McCullough, 2006b), and behavioral experiments (e.g., Bennett-Levy et al., 2004). A systematic review by Nagy and Moore (2017) found that 17 of 24 studies which used interventions that facilitated social interaction (i.e., the *sampling* of social context information) reduced levels of depressive symptoms successfully.

Moreover, the literature from attachment and developmental psychology suggests that the maladaptive attachment styles that can develop as a result of social trauma can trigger pathological patterns in inferring social contexts in depression (Bifulco, Moran, Ball, & Bernazzani, 2002; Bifulco, Moran, Ball, & Lillie, 2002; Brophy et al., 2020; Mikulincer & Shaver, 2012; Murphy & Bates, 1997; Pettem et al., 1993; Sloman, 2003; Spruit et al., 2020). This might be because children develop internal “representations of attachment” with their attachment figures that serve as “templates” for interpreting and predicting behavior in later relationships (Bifulco, Moran, Ball, & Lillie, 2002, p. 60). From this perspective, it seems plausible that depressed individuals with a primarily avoidant attachment style would seldom process social information to avoid being hurt. Vice versa, depressed individuals with a more anxious attachment style may initially exhibit excessive attachment behavior to cope with thwarted needs to belong, but quickly respond with anger or resignation when there is no response (Mikulincer & Shaver, 2012).

As we illustrate below, our computational model of biased social inference in depression offers a new operationalization of various attachment styles in terms of concrete social interactions that extends existing computational approaches (for more details, see Figure 2; Cittern et al., 2018; Constant et al., 2021).

To summarize: individuals with depression seem to exhibit aberrant processing of various types of social information. A multitude of empirical findings suggests biases in how they interpret social cues and problem-solving in interpersonal contexts (Bird et al., 2018; Hames et al., 2013; Kupferberg et al., 2016; Segrin, 1990, 2000; Starr & Davila, 2008; Tse & Bond, 2004; Weightman et al., 2014). These problems often seem to persist and exacerbate negative social beliefs acquired via early-life social trauma or lack of social support. Meanwhile, the underlying mechanisms remain mostly elusive. In the following paragraphs we demonstrate how recent developments in computational neuroscience can be applied to (a) unify a variety of empirical findings on social impairments in depression under a common theoretical framework, (b) improve the operationalization of social impairments in depression, (c) shed light on the elusive computational mechanisms underlying these impairments (e.g., why they develop, are maintained, and manifest in different ways), and (d) functionally link heterogeneous social phenomena in depression to different biases in patients' internal social belief systems.

3. Modelling biased inference of social contexts in depression

3.1 Active inference

In the past decades, emerging computational theories have changed the way/how we think about the mind in health and disease. In cognitive and computational neuroscience, the active inference framework enables us to unify complex processes like perception, learning, action, and decision making under a Bayesian decision-theoretic umbrella (for a synthesis, see Da Costa et al., 2020; for decision-theoretic underpinnings, see Huys et al., 2015). Central to active inference is the notion of the mind as an elaborate inference engine, trying to make sense of its environment from noisy, ambiguous, and partial sensory information (Friston et al., 2017). The "reality" of the external world is not directly accessible to the agent – the state of these *hidden factors* needs to be inferred with the help of an agent's internal, *generative model* (Heins et al., 2022; Smith et al., 2022b). The generative model contains the agent's beliefs about how certain sensory information relates to an environmental factor. In other words, how latent factors or states cause observable (sensory) consequences. This probabilistic model of the world must be continuously updated (i.e., "optimized") by sensory input to ensure the correct inference of hidden states and for maintaining physiological – and psychosocial - homeostasis. It is matched with an agent's *prior* beliefs about certain hidden states or observations (Friston et al., 2016; Smith, Badcock, et al., 2021). Accurate priors and generative models are crucial for the agent's survival and for maintaining physiological homeostasis (for further information, see Da Costa et al., 2020).

A final component of active inference is action. By performing actions or action sequences (i.e., policies), the agent exerts an influence on hidden factors (Heins et al., 2022). Inferring policies allows the agent to select action sequences that lead to desired or preferred sensory information, that is, actions minimize surprise in the long term. In the context of active inference, we can therefore speak of "Predictions, not Commands" (Adams et al., 2013): the motor system executes actions that lead to the expected sensory information. This basic view of sentient behavior can be summarized as choosing courses of action or plans that bring about predicted consequences. This basic idea operates at the level of simple reflexes, through to social narratives. In terms of simple movements, our actions can be viewed as fulfilling top-down predictions of a proprioceptive sort that supply setpoints or equilibrium points (Feldman

& Levin, 1995). This is very reminiscent of ideomotor theory in the 19th-century (Wiese, 2017). When applied to social behaviors and decision-making, this is often cast as planning as inference (Attias, 2003; Botvinick & Toussaint, 2012; Kaplan & Friston, 2018); in other words, committing to those courses of action that minimize the deviation from predicted or expected outcomes.

Mathematically speaking, we can approximate the minimizing of "surprise" by minimizing a quantity named "free energy" (FE). In the context of active inference, minimizing FE should lead to "optimal" perceptions and actions (for further information, see Friston, 2010). More generally, actions or policies are selected that minimize expected free energy (Da Costa et al., 2021). Technically, expected free energy subsumes two kinds of (Bayes) optimality. The first reflects the propensity to bring about preferred outcomes (e.g., being at home), which are just those 'states of being' that characterize the decision maker or person in question. Mathematically, this preference-seeking can be read as maximizing expected value, where value is just the (log) probability of being in a characteristic state. The other part of expected free energy underwrites information-seeking and the resolution of the uncertainty inherent in minimizing expected surprise. Mathematically, this is usually expressed as expected information gain. In short, active inference entails a joint commitment to optimal Bayesian decisions (via maximizing expected value) (Berger, 2011) while, at the same time, complying with the principles of optimum Bayesian design (via maximizing expected information gain) (Lindley, 1956; MacKay, 1992). In effect, this accommodates the exploration-exploitation dilemma by placing epistemic and pragmatic imperatives on the same footing (Schwartenbeck et al., 2019).

As a simplified example, consider a man who needs to infer his physical location in a town. He has no direct access to this information (hidden factor), but needs to infer it from the visual, auditory, and olfactory information his sensory organs take in from his environment.

The agent’s observation model contains his specific beliefs about how certain sensory information relates to a certain neighborhood or location in space (e.g., a sign to a specific area). The agent implicitly forms “hypotheses” about where he is located given his latest sensory information (the *likelihood*). Given such sensory information, the hypothesis with the highest posterior probability wins. Furthermore, his previous actions (i.e. walking down an avenue, then turning right, etc.) provide information about his location in space. Future actions or policies are selected so that the agent’s desired sensory information is acquired. If the agent wants to go home, only those policies will be chosen that will lead to the expected visual, auditory, and olfactory information usually experienced while living at home.

From an evolutionary psychology perspective, active inference constantly involves making a trade-off between seeking new information through action (in this example, exploring the city) and maximizing access to known rewards (returning home quickly) in order to survive in an ever-changing environment (explore vs. exploit dilemma, Gopnik, 2020; Kouvaris et al., 2015). Agents failing to engage in any exploratory behavior who only seek to maximize known rewards “over-fit” their environment, and run the risk of being incapable of adapting to environmental changes (i.e., not finding home when the cityscape changes). Conversely, “under-fitting” agents who engage in a lot of explo

atory behavior and show little reward maximization run the risk of failing to exploit resources and rewards in order to maintain their fitness (i.e., never find home).

In mental health terms, active inference suggests that mental disorders, like depression are the result of biased inference of hidden states of the world (although it is probably not the inference algorithm itself that is biased, but rather the generative model to which it is applied; Schwartenbeck et al., 2015). In this context, Bayesian formulations have proved useful in accounting for different symptom domains of MDD (Chekroud, 2015; Constant et al., 2019;

Smith et al., 2018). Smith and colleagues (2018), for example, suggest that depressive schemas[3] may act as precise, internalized priors in patients with depression. Given a certain set of sensory information, the interpretation, or "hypothesis", with the highest posterior probability is selected to explain the sensory information.[4] In depression, this may be the depressive schema. Moreover, they suggest that the core symptom of psychomotor slowing and loss of energy may result from biased beliefs regarding an action's utility in achieving a desired sensory state (Smith et al., 2018); implicitly, the mind may infer that the cost of a certain action outweighs its utility, thus reducing an agent's motivation for action so that the depressed person feels compelled not to leave their bed. Particularly when processing social information (Suffel et al., 2020), these costs may actually be high, presumably because positive social cues are less expected.

3.2 Active inference of social contexts in depression

Human social interactions are marked by tremendous complexity. In an everyday conversation, noisy verbal- and non-verbal information needs to be considered alongside ambiguous cultural, contextual, and situational factors, all evolving very dynamically and sometimes unpredictably over time. From a Bayesian perspective, the social agent needs to process highly uncertain information and react appropriately in real time, a skill that most humans surprisingly master with relative ease. In depressed individuals, however, interpersonal problems are highly prevalent. We propose here that the distorted interpersonal perception and behaviors in depression are due to various biases in their generative model of the social world's hidden states (i.e., their prior beliefs about how certain sensory information from the social

[3] Here defined as "(…) maladaptive, pessimistic sets of beliefs/expectations that bias perceptual/conceptual interpretations of new sensory input, as well as the subsequent predictions, judgments, and decisions these interpretations inform (…)" (Smith et al., 2018, p. 2)

[4] Please note that this inferential process is often implicit.

environment relates to certain characteristics of social contexts). For example, imagine a party where you only know a few people. A healthy person might go to this party with the prior belief that most of the guests will be kind, and that friendly conversations might ensue (prior belief that the social context will probably be positive). She will interpret the smile of someone sitting next to her as an expression of sympathy (belief that the observation “smile” is highly probable in a positive social context, and less likely in a negative social context) and would expect that she can enhance her social environment’s positive state more by starting a conversation (belief that choosing the “start a conversation” policy would likely transform the current state of the social context into an even more positive one).

In contrast, a depressed person would go to this party with the prior belief that most people will show no interest in her and that probably no one will talk to her (prior belief that the social context is likely to be negative). She will interpret the smile of the person sitting next to her as an expression of pity (belief that the “smile” observation is less probable in a positive social context, and more probable in a negative social context) and would expect that she can improve her social environment’s state by withdrawing socially and by not attracting attention (belief that choosing the policy “social withdrawal” would likely transform the social context’s current state into a less negative one). On the one hand, this could bias further the social context’s inference, since the "social withdrawal" policy enables very little sensory input (e.g., the depressed person gets no information about whether they have common interests or passions with other party guests). This can then lead to this person's internal belief system becoming stabilized (e.g., “immunization”, Kube et al., 2020, or, “learned helplessness” Seligman, 1972). On the other hand, as a kind of self-fulfilling prophecy, this strategy could encourage the very experiences that the depressed person already feared, such as being largely ignored by other people, which could also cause the generative model to become entrenched. As a consequence, the biased inference of social contexts can trigger a self-reinforcing vicious cycle in depression that can result in increasing discord between the social environment and interpersonal behavior.

Practitioners know that abnormalities in social perception and social behavior can be very heterogeneous in depressed individuals (e.g., more avoidant vs. more hostile/suspicious interpersonal styles). We believe this can be attributed to different forms of biases within these patients' given generative models.[5] The flexibility inherent to the active inference approach enables us to characterize these different "phenotypes" of pathological inference in depression as well as more "healthy" forms of social inference within the same framework (see Figure 2). We elaborate on this idea in the following sections.

3.3 Creating a formal model

Clinical applications of partially observable Markov decision process (POMDP) models have been described (e.g., Smith et al., 2020, 2022). We describe here the key concepts and parameters briefly again and apply them to the active inference of social contexts during brief social interactions. Agents have beliefs about the state of a social context (s) and about their possible action sequences in that context (π). Agents make observations (o) mediated by their senses. Furthermore, agents have beliefs about the probability of being in a particular social context (s_T) when making a particular observation (o_T). This is the agent's observation model, which is represented by a probability distribution $p(o_T|s_T)$, containing information about how

[5] In fact, any social experience and behavior must be expressed as deviations of some form in the probability distributions that constitute the generative model. This truism is known as the complete class theorem (Brown, 1981; Wald, 1947), which says that for any pair of behaviors and value functions, there exists some prior beliefs that render the behavior Bayes optimal. In short, the ideal Bayesian assumptions inherent in active inference (please see above) means that one can describe any behavior in terms of the agent's prior beliefs. This is one of the main motivations for computational phenotyping of the sort that characterizes a patient's choice behavior in terms of their underlying prior beliefs (Smith, Kirlic, et al., 2021).

probable it is that states will bring about certain observations. Further, the agent entertains beliefs about the probability of transitioning from one state to another, possibly influenced by their own actions. This is represented by a probability distribution $p(s_{T+1}|s_{T}, \pi)$. Finally, agents may possess initial beliefs about the hidden nature of a social context $p(s_1)$.

[Figure 1]

From the perspective of this model, social experience and behavior in depression might be attributable to deviations from healthy participants in the probability distributions mentioned above. As previously discussed, there is evidence that patients with depression tend to have more negative beliefs about the hidden nature of social contexts. This could be modelled as a prior over the initial hidden social state ($p(s_1)$) – where a negative, hostile state is assigned higher prior probability than a benevolent one at the onset of the social interaction (e.g. „they won‘t like me“). Patients may furthermore reveal biased beliefs about how likely certain observations are to occur in a given social context. For example, a patient may carry high expectations that a stranger will initiate a conversation with them if they are liked, otherwise inferring a hostile social state, i.e. that he or she is not liked by the other („if they liked me, they would approach me and start a conversation“). This factor concerns the patient's observation model $p(o_T|s_T)$. Patients may differ still more from non-depressed individuals in their pessimistic beliefs about how likely certain action sequences will alter the social context – socially learned helplessness would dictate that the patient's actions would have no measurable consequences in the patient's prediction („it wouldn‘t change anything if I smile at them“). This relates to transition-probabilities $p(s_{T+1}|s_{T}, \pi)$.

Depression is a very heterogeneous disorder which active inference could account for by allowing for variations within the generative model. This may vary from patient to patient, depending on the individual's learning history. For example, depressed individuals whose early attachment figures were highly impulsive and ambivalent may reveal great uncertainty about their present social context (which would correspond to imprecise priors on initial hidden states). In contrast, depressed individuals affected by devaluing and punishing attachment figures would predict with great certainty that they find themselves in a negative social context when encountering a social situation in the present (this would correspond, in turn, to very precise priors of negative hidden states). In addition to changes in parameter $p(s_1)$, changes in our model's other parameters lead to potentially additional problems in the inference of social contexts. For a brief outline, see Figure 2.

[Figure 2]

5. Implications and future research

Our approach enables a new formalization of the active inference of social context characteristics in mentally healthy and ill individuals, which allows researchers to functionally link heterogeneous social phenomena in depression to different biases in patient's internal social belief systems. In consequence, it offers new perspectives for both etiological research in depression as well as improved diagnostics and treatment. Knowing *how* social trauma or the lack of social support affects certain parameters of the generative model and *which* parameters (or parameter configurations) are predestined for certain social experiences and behaviors has profound implications for prevention and treatment-selection, since treatment strategies affect the patient's generative model's parameters *differently*. To stay with the above-mentioned example: Individuals with "too low" and barely deviant scores on parameter $p(s_1)$ (i.e., with great uncertainty about which social context they will enter), are likely to benefit most from interventions that reduce social safety behaviors and provide consistent reinforcement schedule within the therapeutic relationship, since these interventions foster beliefs in the predictability of social contexts in the long term. Conversely, individuals with a strong negative prior (i.e., who are deeply convinced they will enter a negative social context in the future) are more likely to benefit from reparenting techniques and interpersonal discrimination training, as they improve social context discrimination. This example illustrates that different parameter specifications in the generative model can lead to various treatment implications. Moreover, it illustrates how the treatment of depressive disorders can be individualized by our conceptualization, even when different distortions in the same parameter are involved.

In a similar way, the approach we suggest could deepen our understanding of the neural implementation of related processes. While the literature has yielded quite heterogeneous findings on the neural correlates of depression (Gray et al., 2020) a better description of impairments on a computational level might reduce variability across studies and reveal more suitable approaches for tailored interventions (e.g., brain stimulation, Brunoni et al., 2017).

Finally, our conceptualization of social biases in depression sheds light on the issue of treatment-resistant depression (Schramm et al., 2020). It seems plausible that the interventions these patients undergo often fail to address the “right” parameters in the respective generative models. For example, no matter how many positive social experiences a depressed patient may have with other people during therapy, if he or she believes that positive social cues are unrelated to the social context’s actual nature (concerns the observational model $p(o_T|s_T)$), then these experiences have no impact on the generative model.

From an etiological and developmental perspective, research should examine further which social experiences affect model parameters over the life span. This would also constitute a thorough test of complex interpersonal theories on the development of depressive disorders, such as those assumed in attachment theory, schema therapy or the Cognitive-Behavioral Analysis System of Psychotherapy (CBASP). In this context, we assume that our model can also be applied to interpersonal experience and behavior in other mental disorders like personality disorders or social phobia (e.g., Mikulincer & Shaver, 2012).

Furthermore, research should focus on accurately diagnosing patient’s generative models, linking them to concrete treatment recommendations. From the practitioner's perspective, the relevant questions are (a) how to diagnose generative models in practice and (b) which interventions are most suitable to influence certain model parameters. These recommendations could be followed in turn to better tailor the treatment of depressive disorders. In this context, we suggest adapting experimental paradigms from social psychology (e.g., Wirth, 2016) or game theory (e.g., Wang et al., 2015), as these paradigms require social decisions based on inferences about social contexts and can be easily applied in practice as well as formalized in terms of the model above (cf. Smith et al., 2022). For example, social decision making in cooperative games that require inferences about social contexts (e.g., evaluating a co-players "trustworthiness"; cf. Wang et al., 2015) could be used to estimate the generative model of depressed patients and systematically compare it to the model of healthy controls. In clinical practice, the availability of behavioral data throughout the diagnostic process could provide clinicians with a concrete idea of their patient's implicit internal models of their social environment. This could be exploited, again, to individualize therapies.

Our proposed research agenda is in line with a broader trend away from fixed clinical categories towards a more multidimensional perspective on diagnosing mental disorders (Insel et al., 2010). Multidimensional approaches can account for heterogeneity in patient populations and shed light on shared mechanisms in what have been assumed to be distinct disorders. In the future, computational markers of depression may prove to be beneficial in arriving at a precise description and selecting optimal treatment strategies. These formalizations could objectify the diagnostics of mental disorders in general, and strengthen research on shared mechanisms.

From a methodological perspective, future research should – as the first step – specify several prototypes of generative models that describe the active inference of social contexts in healthy vs. clinical samples, as well as between different "phenotypes" of the disorders under investigation. These models should be guided by theories regarding different etiologies of each disorder. As the second step, these generative models could serve to simulate behavior, and then matched to empirical data from healthy and clinical samples. For example, parameter estimates could be obtained from (social) decision behavior (Smith et al., 2022b) in standard experimental paradigms (e.g., game theoretical tasks, Wang et al., 2015). In the final step, researchers should identify those generative models that yield the best model fit and work on interventions to optimize their parameters with respect to mental health.

Finally, another important step would be to develop a model applied to describe the interaction of multiple agents with different generative POMDP models. This would enable researchers to study dynamic interaction patterns in health and disease, and their effects on the corresponding generative model of the agents involved.

6. Conclusion

There is broad consensus that social functioning, i.e., the anticipation, perception, and processing of social information, as well as social behavior, is impaired in depression. We attribute these impairments primarily to deviations in patients' internal models of their social environment, leading to biased inferences about social contexts in the present. Formalizing these biases within the active inference framework and operationalizing them in the context of experimental paradigms that involve social inferences (e.g., socioeconomic decision tasks) can shed light on the wide heterogeneity of depressive symptomatology and provide insights into the underlying mechanisms and configurations of patients' internal models of their social world. Determining the generative model of patients in social contexts would improve both diagnostics and research, as well as facilitate treatment selection for mental disorders.

Authors Contributions

Conceptualization: L. Kirchner, A.-L. Eckert, M. Berg, D. Endres, and W. Rief; Methodology: L. Kirchner, A.-L. Eckert, M. Berg, D. Endres, B. Straube; Writing - Original Draft: L. Kirchner; Writing - Review & Editing: L. Kirchner, A.-L. Eckert, M. Berg, D. Endres, B. Straube, W. Rief; Resources: A.-L. Eckert, L. Kirchner, D. Endres; Visualization: L. Kirchner, A.-L. Eckert. Funding Acquisition: This work was funded by the Hessian Ministry of Higher Education, Research, Science, and the Arts as part of the cluster initiative „The Adaptive Mind“.

Acknowledgments

I would like to thank all the collaborators of the cluster initiative "The Adaptive Mind" (TAM) for the productive discussions that made this project possible.

Funding

This work was funded by the Hessian Ministry of Higher Education, Research, Science, and the Arts as part of the cluster initiative „The Adaptive Mind“.

Conflict of interest

This work was conducted as part of the cluster initiative “The Adaptive Mind” (TAM) which brings together scientists from experimental psychology, clinical psychology, and artificial intelligence to improve the understanding of how the human mind successfully adapts to changing conditions.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Attias, H. (2003). Planning by probabilistic inference. In C. M. Bishop & B. J. Frey (Eds.), *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*. Retrieved from https://proceedings.mlr.press/r4/attias03a.html

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: an evolutionary systems theory. *Trends in Cognitive Sciences*, *21*(3), 182–194. https://doi.org/10.1016/j.tics.2017.01.005

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin*, *112*(3), 461–484. https://doi.org/10.1037/0033-2909.112.3.461

Beck, A. T. (1963). Thinking and depression: I. idiosyncratic content and cognitive distortions. *Archives of General Psychiatry*, *9*(4), 324–333. https://doi.org/10.1001/ARCHPSYC.1963.01720160014002

Beck, A. T. (1964). Thinking and depression: II. theory and therapy. *Archives of General Psychiatry*, *10*(6), 561. https://doi.org/10.1001/archpsyc.1964.01720240015003

Beeson, C. M. L., Brittain, H., & Vaillancourt, T. (2020). The temporal precedence of peer rejection, rejection sensitivity, depression, and aggression across adolescence. *Child Psychiatry & Human Development*, *51*(5), 781–791. https://doi.org/10.1007/s10578-020-01008-2

Bellack, A. S., Hersen, M., & Himmelhoch, J. M. (1996). Social skills training for depression. In *Sourcebook of Psychological Treatment Manuals for Adult Disorders* (pp. 179–200). Springer US. https://doi.org/10.1007/978-1-4899-1528-3_5

Bennett-Levy, J., Butler, G., Fennell, M., Hackmann, A., Mueller, M., & Westbrook, D. (2004). Oxford Guide to Behavioural Experiments in Cognitive Therapy. In J. Bennett-Levy, G. Butler, M. Fennell, A. Hackmann, M. Mueller, & D. Westbrook (Eds.), *Oxford Guide to Behavioural Experiments in Cognitive Therapy*. Oxford University Press. https://doi.org/10.1093/med:psych/9780198529163.001.0001

Berger, J. O. (2011). *Statistical decision theory and Bayesian analysis*. New York: Springer.

Bifulco, A., Moran, P. M., Ball, C., & Bernazzani, O. (2002). Adult attachment style. I: its relationship to clinical depression. *Social Psychiatry and Psychiatric Epidemiology*, *37*(2), 50–59. https://doi.org/10.1007/s127-002-8215-0

Bifulco, A., Moran, P. M., Ball, C., & Lillie, A. (2002). Adult attachment style. II: its relationship to psychosocial depressive-vulnerability. *Social Psychiatry and Psychiatric Epidemiology*, *37*(2), 60–67. https://doi.org/10.1007/S127-002-8216-X

Bird, T., Tarsia, M., & Schwannauer, M. (2018). Interpersonal styles in major and chronic depression: a systematic review and meta-analysis. *Journal of Affective Disorders*, *239*, 93–101. https://doi.org/10.1016/j.jad.2018.05.057

Bishop, A., Younan, R., Low, J., & Pilkington, P. D. (2022). Early maladaptive schemas and depression in adulthood: a systematic review and meta-analysis. *Clinical Psychology & Psychotherapy*, *29*(1), 111–130. https://doi.org/10.1002/cpp.2630

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485–488. https://doi.org/10.1016/j.tics.2012.08.006

Brophy, K., Brähler, E., Hinz, A., Schmidt, S., & Körner, A. (2020). The role of self-compassion in the relationship between attachment, depression, and quality of life. *Journal of Affective Disorders*, *260*, 45–52. https://doi.org/10.1016/j.jad.2019.08.066

Brown, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *The Annals of Statistics*, *9*(6), 1289–1300. https://doi.org/10.1214/aos/1176345645

Brunoni, A. R., Moffa, A. H., Sampaio-Junior, B., Borrione, L., Moreno, M. L., Fernandes, R. A., Veronezi, B. P., Nogueira, B. S., Aparicio, L. V. M., Razza, L. B., Chamorro, R., Tort, L. C., Fraguas, R., Lotufo, P. A., Gattaz, W. F., Fregni, F., & Benseñor, I. M. (2017). Trial of electrical direct-current therapy versus escitalopram for depression. *New England Journal of Medicine*, *376*(26), 2523–2533. https://doi.org/10.1056/NEJMoa1612999

Chekroud, A. M. (2015). Unifying treatments for depression: an application of the Free Energy Principle. *Frontiers in Psychology*, *6*(FEB), 153. https://doi.org/10.3389/FPSYG.2015.00153/BIBTEX

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: a review of computational research. *Neuroscience & Biobehavioral Reviews*, *55*, 247–267. https://doi.org/10.1016/j.neubiorev.2015.05.005

Cittern, D., Nolte, T., Friston, K., & Edalat, A. (2018). Intrinsic and extrinsic motivators of attachment under active inference. *PLOS ONE*, *13*(4), e0193955. https://doi.org/10.1371/journal.pone.0193955

Constant, A., Hesp, C., Davey, C. G., Friston, K. J., & Badcock, P. B. (2021). Why depressed mood is adaptive: a numerical proof of principle for an evolutionary systems theory of depression. *Computational Psychiatry*, *5*(1), 60–80. https://doi.org/10.5334/cpsy.70

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., & Friston, K. (2019). Regimes of expectations: an active inference model of social conformity and human decision making. *Frontiers in Psychology*, *10*(MAR), 679. https://doi.org/10.3389/FPSYG.2019.00679/BIBTEX

Coyne, J. C. (1976). Toward an interactional description of depression. *Psychiatry*, *39*(1), 28–40. https://doi.org/10.1080/00332747.1976.11023874

Cruwys, T., Haslam, S. A., Dingle, G. A., Haslam, C., & Jetten, J. (2014). Depression and social identity: an integrative review. *Personality and Social Psychology Review*, *18*(3), 215–238. https://doi.org/10.1177/1088868314523839

Cuijpers, P., Muñoz, R. F., Clarke, G. N., & Lewinsohn, P. M. (2009). Psychoeducational treatment and prevention of depression: the "coping with depression" course thirty years later. *Clinical Psychology Review*, *29*(5), 449–458. https://doi.org/10.1016/j.cpr.2009.04.005

D'Iuso, D. A., Dobson, K. S., Beaulieu, L., & Drapeau, M. (2018). Coping and interpersonal functioning in depression. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *50*(4), 248–255. https://doi.org/10.1037/cbs0000112

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: a synthesis. *Journal of Mathematical Psychology*, *99*, 102447. https://doi.org/10.1016/j.jmp.2020.102447

Da Costa, L., Parr, T., Sengupta, B., & Friston, K. (2021). Neural dynamics under active inference: plausibility and efficiency of information processing. *Entropy*, *23*(4), 454. https://doi.org/10.3390/e23040454

Dobson, K. S., Quigley, L., & Dozois, D. J. A. (2014). Toward an integration of interpersonal risk models of depression and cognitive-behaviour therapy. *Australian Psychologist*, *49*(6), 328–336. https://doi.org/10.1111/ap.12079

Everaert, J., Podina, I. R., & Koster, E. H. W. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review*, *58*, 33–48. https://doi.org/10.1016/j.cpr.2017.09.005

Evraire, L. E., & Dozois, D. J. A. (2011). An integrative model of excessive reassurance seeking and negative feedback seeking in the development and maintenance of depression. *Clinical Psychology Review*, *31*(8), 1291–1303. https://doi.org/10.1016/j.cpr.2011.07.014

Feldman, A. G., & Levin, M. F. (1995). The origin and use of positional frames of reference in motor control. *Behavioral and Brain Sciences*, *18*(4), 723–744. https://doi.org/10.1017/S0140525X0004070X

Fernández-Theoduloz, G., Paz, V., Nicolaisen-Sobesky, E., Pérez, A., Buunk, A. P., Cabana, Á., & Gradin, V. B. (2019). Social avoidance in depression: a study using a social decision-making task. *Journal of Abnormal Psychology*, *128*(3), 234–244. https://doi.org/10.1037/abn0000415

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, *68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. *Neural Computation*, *29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Gao, S., Assink, M., Cipriani, A., & Lin, K. (2017). Associations between rejection sensitivity and mental health outcomes: a meta-analytic review. *Clinical Psychology Review*, *57*, 59–74. https://doi.org/10.1016/j.cpr.2017.08.007

Gariépy, G., Honkaniemi, H., & Quesnel-Vallée, A. (2016). Social support and protection from depression: systematic review of current findings in Western countries. *British*

Journal of Psychiatry, *209*(4), 284–293. https://doi.org/10.1192/bjp.bp.115.169094

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1803), 20190502. https://doi.org/10.1098/rstb.2019.0502

Gray, J. P., Müller, V. I., Eickhoff, S. B., & Fox, P. T. (2020). Multimodal abnormalities of brain structure and function in major depressive disorder: a meta-analysis of neuroimaging studies. *American Journal of Psychiatry*, *177*(5), 422–434. https://doi.org/10.1176/appi.ajp.2019.19050560

Hames, J. L., Hagan, C. R., & Joiner, T. E. (2013). Interpersonal processes in depression. *Annual Review of Clinical Psychology*, *9*, 355–377. https://doi.org/10.1146/annurev-clinpsy-050212-185553

Heins, C., Millidge, B., Demekas, D., Klein, B., Friston, K., Couzin, I. D., & Tschantz, A. (2022). pymdp: a Python library for active inference in discrete state spaces. *Journal of Open Source Software*, *7*(73), 4098. https://doi.org/10.21105/joss.04098

Huys, Q. J. M., Guitart-Masip, M., Dolan, R. J., & Dayan, P. (2015). Decision-theoretic psychiatry. *Clinical Psychological Science*, *3*(3), 400–421. https://doi.org/10.1177/2167702614562040

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, *112*(4), 323–343. https://doi.org/10.1007/s00422-018-0753-2

Kasch, K. L., Rottenberg, J., Arnow, B. A., & Gotlib, I. H. (2002). Behavioral activation and inhibition systems and the severity and course of depression. *Journal of Abnormal Psychology*, *111*(4), 589–597. https://doi.org/10.1037/0021-843X.111.4.589

Kirchner, L., Eckert, A., & Berg, M. (2022). From broken models to treatment selection: Active inference as a tool to guide clinical research and practice. *Clinical Psychology in Europe*, *4*(2), 0–4. https://doi.org/10.32872/cpe.9697

Kirchner, L., Schummer, S. E., Krug, H., Kube, T., & Rief, W. (2022). How social rejection expectations and depressive symptoms bi-directionally predict each other – A cross-lagged panel analysis. *Psychology and Psychotherapy: Theory, Research and Practice*, *95*(2), 477–492. https://doi.org/10.1111/papt.12383

Klerman, G. L., & Weissman, M. M. (1994). *Interpersonal psychotherapy of depression: a brief, focused, specific strategy*. Retrieved from https://rowman.com/ISBN/9781568213507/Interpersonal-Psychotherapy-of-Depression-A-Brief-Focused-Specific-Strategy

Kouvaris, K., Clune, J., Kounios, L., Brede, M., & Watson, R. A. (2015). How evolution learns to generalise: principles of under-fitting, over-fitting and induction in the evolution of developmental organisation. *Nature*, *388*, 539–547. https://doi.org/10.48550/arXiv.1508.06854

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: a predictive processing perspective. *Biological Psychiatry*, *87*(5), 388–398. https://doi.org/10.1016/j.biopsych.2019.07.017

Kupferberg, A., Bicks, L., & Hasler, G. (2016). Social functioning in major depressive disorder. *Neuroscience and Biobehavioral Reviews*, *69*, 313–332. https://doi.org/10.1016/j.neubiorev.2016.07.002

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The*

Annals of Mathematical Statistics, *27*(4), 986–1005. https://doi.org/10.1214/aoms/1177728069

Lipsitz, J. D., & Markowitz, J. C. (2013). Mechanisms of change in interpersonal therapy (IPT). *Clinical Psychology Review*, *33*(8), 1134–1147. https://doi.org/10.1016/j.cpr.2013.09.002

MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation*, *4*(4), 590–604. https://doi.org/10.1162/neco.1992.4.4.590

Mccullough, J. P. (2006). *Treating chronic depression with disciplined personal involvement: cognitive behavioral analysis system of psychotherapy (CBASP)*. Springer.

McCullough, J. P. (2003). Treatment for chronic depression using cognitive behavioral analysis system of psychotherapy (CBASP). *Journal of Clinical Psychology*, *59*(8), 833–846. https://doi.org/10.1002/jclp.10176

McCullough, J. P. (2006). Healing interpersonal trauma using the interpersonal discrimination exercise. In *Treating Chronic Depression with Disciplined Personal Involvement* (pp. 123–159). Springer US. https://doi.org/10.1007/978-0-387-31066-4_6

McCullough, J., Schramm, E., & Penberthy, J. K. (2014). CBASP as a distinctive treatment for persistent depressive disorder. In *CBASP as a Distinctive Treatment for Persistent Depressive Disorder*. Routledge. https://doi.org/10.4324/9781315743196

Mikulincer, M., & Shaver, P. R. (2012). An attachment perspective on psychopathology. *World Psychiatry*, *11*(1), 11–15. https://doi.org/10.1016/j.wpsyc.2012.01.003

Muñoz, R. F., Ghosh Ippen, C., Rao, S., Le, H.-N., & Valdes Dwyer, E. (2000). *Manual for group cognitive-behavioral therapy of major depression*. Retrieved from https://i4health.paloaltou.edu/downloads/CBT_Instructor_English.pdf

Murphy, B., & Bates, G. W. (1997). Adult attachment style and vulnerability to depression. *Personality and Individual Differences*, *22*(6), 835–844. https://doi.org/10.1016/S0191-8869(96)00277-2

Nagy, E., & Moore, S. (2017). Social interventions: an effective approach to reduce adult depression? *Journal of Affective Disorders*, *218*, 131–152. https://doi.org/10.1016/j.jad.2017.04.043

Nanni, V., Uher, R., & Danese, A. (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *American Journal of Psychiatry*, *169*(2), 141–151. https://doi.org/10.1176/appi.ajp.2011.11020335

Nestor, B. A., Sutherland, S., & Garber, J. (2022). Theory of mind performance in depression: a meta-analysis. *Journal of Affective Disorders*, *303*, 233–244. https://doi.org/10.1016/j.jad.2022.02.028

Panitz, C., Endres, D., Buchholz, M., Khosrowtaj, Z., Sperl, M. F. J., Mueller, E. M., Schubö, A., Schütz, A. C., Teige-Mocigemba, S., & Pinquart, M. (2021). A revised framework for the investigation of expectation update versus maintenance in the context of expectation violations: the ViolEx 2.0 model. *Frontiers in Psychology*, *12*(November). https://doi.org/10.3389/fpsyg.2021.726432

Pettem, O., West, M., Mahoney, A., & Keller, A. (1993). Depression and attachment problems. *Journal of Psychiatry & Neuroscience : JPN*, *18*(2), 78–81. https://pubmed.ncbi.nlm.nih.gov/8461287/

Platt, B., Kadosh, K. C., & Lau, J. Y. F. (2013). The role of peer rejection in adolescent depression. *Depression and Anxiety*, *30*(9), 809–821. https://doi.org/10.1002/da.22120

Porcelli, S., Van Der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., Mogavero, F., Lobo, A., Olivera, F. J., Lobo, E., Posadas, M., Dukart, J., Kozak, R., Arce, E., Ikram, A., Vorstman, J., Bilderbeck, A., Saris, I., Kas, M. J., & Serretti, A. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews*, *97*, 10–33. https://doi.org/10.1016/j.neubiorev.2018.09.012

Rantala, M. J., Luoto, S., Krams, I., & Karlsson, H. (2018). Depression subtyping based on evolutionary psychiatry: proximate mechanisms and ultimate functions. *Brain, Behavior, and Immunity*, *69*, 603–617. https://doi.org/10.1016/j.bbi.2017.10.012

Renner, F., Arntz, A., Leeuw, I., & Huibers, M. (2013). Treatment for chronic depression using schema therapy. *Clinical Psychology: Science and Practice*, *20*(2), 166–180. https://doi.org/10.1111/cpsp.12032

Rief, W., & Joormann, J. (2019). Revisiting the cognitive model of depression: the role of expectations. *Clinical Psychology in Europe*, *1*(1), 1–19. https://doi.org/10.32872/cpe.v1i1.32605

Roediger, E., Stevens, B. A., & Brockman, R. (2018). *Contextual schema therapy: An integrative approach to personality disorders, emotional dysregulation, and interpersonal functioning. - PsycNET*. New Harbinger Publications. Retrieved from https://psycnet.apa.org/record/2018-30793-000

Roepke, A. M., & Seligman, M. E. P. (2016). Depression and prospection. *British Journal of Clinical Psychology*, *55*(1), 23–48. https://doi.org/10.1111/bjc.12087

Safra, L., Chevallier, C., & Palminteri, S. (n.d.). *Social information impairs reward learning in depressive subjects: behavioral and computational characterization*. https://doi.org/10.1101/378281

Safra, L., Chevallier, C., & Palminteri, S. (2019). Depressive symptoms are associated with blunted reward learning in social contexts. *PLOS Computational Biology*, *15*(7), e1007224. https://doi.org/10.1371/journal.pcbi.1007224

Schramm, E., Klein, D. N., Elsaesser, M., Furukawa, T. A., & Domschke, K. (2020). Review of dysthymia and persistent depressive disorder: history, correlates, and clinical implications. *The Lancet Psychiatry*, *7*(9), 801–812. https://doi.org/10.1016/S2215-0366(20)30099-7

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., & Friston, K. (2015). Optimal inference with suboptimal models: addiction and active Bayesian inference. *Medical Hypotheses*, *84*(2), 109–117. https://doi.org/10.1016/j.mehy.2014.12.007

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *ELife*, *8*, e41703. https://doi.org/10.7554/eLife.41703

Segrin, C. (1990). A meta-analytic review of social skill deficits in depression. *Communication Monographs*, *57*(4), 292–308. https://doi.org/10.1080/03637759009376204

Segrin, C. (2000). Social skills deficits associated with depression. *Clinical Psychology Review*, *20*(3), 379–403. https://doi.org/10.1016/S0272-7358(98)00104-4

Seidel, E.-M., Habel, U., Finkelmeyer, A., Schneider, F., Gur, R. C., & Derntl, B. (2010). Implicit and explicit behavioral tendencies in male and female depression. *Psychiatry Research*, *177*(1–2), 124–130. https://doi.org/10.1016/j.psychres.2010.02.001

Seligman, M. E. P. (1972). Learned helplessness. *Annual Review of Medicine*, *23*(1), 407–412. https://doi.org/10.1146/annurev.me.23.020172.002203

Sloman, L. (2003). Evolved mechanisms in depression: the role and interaction of attachment and social rank in depression. *Journal of Affective Disorders*, *74*(2), 107–121. https://doi.org/10.1016/S0165-0327(02)00116-7

Smith, R., Alkozei, A., Killgore, W. D. S., & Lane, R. D. (2018). Nested positive feedback loops in the maintenance of major depression: an integration and extension of previous models. *Brain, Behavior, and Immunity*, *67*, 374–397. https://doi.org/10.1016/J.BBI.2017.09.011

Smith, R., Badcock, P., & Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences*, *75*(1), 3–13. https://doi.org/10.1111/pcn.13138

Smith, R., Friston, K. J., & Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, *107*, 102632. https://doi.org/10.1016/j.jmp.2021.102632

Smith, R., Kirlic, N., Stewart, J. L., Touthang, J., Kuplicki, R., Khalsa, S. S., Feinstein, J., Paulus, M. P., & Aupperle, R. L. (2021). Greater decision uncertainty characterizes a transdiagnostic patient sample during approach-avoidance conflict: a computational modelling approach. *Journal of Psychiatry and Neuroscience*, *46*(1), E74–E87. https://doi.org/10.1503/jpn.200032

Smith, R., Schwartenbeck, P., Stewart, J. L., Kuplicki, R., Ekhtiari, H., & Paulus, M. P. (2020). Imprecise action selection in substance use disorder: evidence for active learning impairments when solving the explore-exploit dilemma. *Drug and Alcohol Dependence*, *215*(April). https://doi.org/10.1016/j.drugalcdep.2020.108208

Spruit, A., Goos, L., Weenink, N., Rodenburg, R., Niemeyer, H., Stams, G. J., & Colonnesi, C. (2020). The relation between attachment and depression in children and adolescents: a multilevel meta-analysis. *Clinical Child and Family Psychology Review*, *23*(1), 54–69. https://doi.org/10.1007/s10567-019-00299-9

Starr, L. R., & Davila, J. (2008). Excessive reassurance seeking, depression, and interpersonal rejection: a meta-analytic review. *Journal of Abnormal Psychology*, *117*(4), 762–775. https://doi.org/10.1037/a0013866

Stellrecht, N. E., Jr., T. E. J., & Rudd, M. D. (2006). Responding to and treating negative interpersonal processes in suicidal depression. *Journal of Clinical Psychology*, *62*(9), 1129–1140. https://doi.org/10.1002/jclp.20298

Suffel, A., Nagels, A., Steines, M., Kircher, T., & Straube, B. (2020). Feeling addressed - the neural processing of social communicative cues in patients with major depression. *Human Brain Mapping*, *41*(13), 3541–3554. https://doi.org/10.1002/hbm.25027

Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: an integrative model. *Clinical Psychology Review*, *31*(7), 1156–1168. https://doi.org/10.1016/j.cpr.2011.07.007

Tse, W. S., & Bond, A. J. (2004). The impact of depression on social skills. *Journal of Nervous & Mental Disease*, *192*(4), 260–268. https://doi.org/10.1097/01.nmd.0000120884.60002.2b

Uphoff, E., Ekers, D., Robertson, L., Dawson, S., Sanger, E., South, E., Samaan, Z., Richards, D., Meader, N., & Churchill, R. (2020). Behavioural activation therapy for depression in adults. *Cochrane Database of Systematic Reviews*, *2020*(7), CD013305. https://doi.org/10.1002/14651858.CD013305.PUB2/FULL

Wald, A. (1947). An essentially complete class of admissible decision functions. *The Annals of Mathematical Statistics*, *18*(4), 549–555. https://doi.org/10.1214/aoms/1177730345

Wang, Y., Yang, L. Q., Li, S., & Zhou, Y. (2015). Game theory paradigm: a new tool for investigating social dysfunction in major depressive disorders. *Frontiers in Psychiatry*, *6*(SEP), 128. https://doi.org/10.3389/FPSYT.2015.00128/BIBTEX

Watson, P. J., & Andrews, P. W. (2002). Toward a revised evolutionary adaptationist analysis of depression: the social navigation hypothesis. *Journal of Affective Disorders*, *72*(1), 1–14. https://doi.org/10.1016/S0165-0327(01)00459-1

Weightman, M. J., Air, T. M., & Baune, B. T. (2014). A review of the role of social cognition in major depressive disorder. *Frontiers in Psychiatry*, *5*(NOV), 179. https://doi.org/10.3389/FPSYT.2014.00179/BIBTEX

Wiese, W. (2017). Action is enabled by systematic misrepresentations. *Erkenntnis*, *82*(6), 1233–1252. https://doi.org/10.1007/s10670-016-9867-x

Wirth, J. H. (2016). Methods for investigating social exclusion. In *Social Exclusion* (pp. 25–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-33033-4_2

Zimmer-Gembeck, M. J. (2016). Peer rejection, victimization, and relational self-system processes in adolescence: toward a transactional model of stress, coping, and developing sensitivities. *Child Development Perspectives*, *10*(2), 122–127. https://doi.org/10.1111/cdep.12174

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Attias, H. (2003). Planning by probabilistic inference. In C. M. Bishop & B. J. Frey (Eds.), *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics* (Vol. R4, pp. 9–16). PMLR. https://proceedings.mlr.press/r4/attias03a.html

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: an evolutionary systems theory. *Trends in Cognitive Sciences*, *21*(3), 182–194. https://doi.org/10.1016/j.tics.2017.01.005

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin*, *112*(3), 461–484. https://doi.org/10.1037/0033-2909.112.3.461

Beck, A. T. (1963). Thinking and depression: I. idiosyncratic content and cognitive distortions. *Archives of General Psychiatry*, *9*(4), 324–333. https://doi.org/10.1001/ARCHPSYC.1963.01720160014002

Beck, A. T. (1964). Thinking and depression: II. theory and therapy. *Archives of General Psychiatry*, *10*(6), 561. https://doi.org/10.1001/archpsyc.1964.01720240015003

Beeson, C. M. L., Brittain, H., & Vaillancourt, T. (2020). The Temporal Precedence of Peer Rejection, Rejection Sensitivity, Depression, and Aggression Across Adolescence. *Child Psychiatry & Human Development*, *51*(5), 781–791. https://doi.org/10.1007/s10578-020-01008-2

Bellack, A. S., Hersen, M., & Himmelhoch, J. M. (1996). Social skills training for depression. In *Sourcebook of Psychological Treatment Manuals for Adult Disorders* (pp. 179–200). Springer US. https://doi.org/10.1007/978-1-4899-1528-3_5

Bennett-Levy, J., Butler, G., Fennell, M., Hackmann, A., Mueller, M., & Westbrook, D. (2004). Oxford Guide to Behavioural Experiments in Cognitive Therapy. In J. Bennett-Levy, G. Butler, M. Fennell, A. Hackmann, M. Mueller, & D. Westbrook (Eds.), *Oxford Guide to Behavioural Experiments in Cognitive Therapy*. Oxford University Press. https://doi.org/10.1093/med:psych/9780198529163.001.0001

Berger, J. O. (2011). *Statistical decision theory and Bayesian analysis*. Springer.

Bifulco, A., Moran, P. M., Ball, C., & Bernazzani, O. (2002). Adult attachment style. I: its relationship to clinical depression. *Social Psychiatry and Psychiatric Epidemiology*, *37*(2), 50–59. https://doi.org/10.1007/s127-002-8215-0

Bifulco, A., Moran, P. M., Ball, C., & Lillie, A. (2002). Adult attachment style. II: its relationship to psychosocial depressive-vulnerability. *Social Psychiatry and Psychiatric Epidemiology*, *37*(2), 60–67. https://doi.org/10.1007/S127-002-8216-X

Bird, T., Tarsia, M., & Schwannauer, M. (2018). Interpersonal styles in major and chronic depression: a systematic review and meta-analysis. *Journal of Affective Disorders*, *239*, 93–101. https://doi.org/10.1016/j.jad.2018.05.057

Bishop, A., Younan, R., Low, J., & Pilkington, P. D. (2022). Early maladaptive schemas and depression in adulthood: a systematic review and meta-analysis. *Clinical Psychology & Psychotherapy*, *29*(1), 111–130. https://doi.org/10.1002/cpp.2630

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485–488. https://doi.org/10.1016/j.tics.2012.08.006

Brophy, K., Brähler, E., Hinz, A., Schmidt, S., & Körner, A. (2020). The role of self-compassion in the relationship between attachment, depression, and quality of life. *Journal of Affective Disorders*, *260*, 45–52. https://doi.org/10.1016/j.jad.2019.08.066

Brown, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *The Annals of Statistics*, *9*(6), 1289–1300. https://doi.org/10.1214/aos/1176345645

Brunoni, A. R., Moffa, A. H., Sampaio-Junior, B., Borrione, L., Moreno, M. L., Fernandes, R. A., Veronezi, B. P., Nogueira, B. S., Aparicio, L. V. M., Razza, L. B., Chamorro, R., Tort, L. C., Fraguas, R., Lotufo, P. A., Gattaz, W. F., Fregni, F., & Benseñor, I. M. (2017). Trial of electrical direct-current therapy versus escitalopram for depression. *New England Journal of Medicine*, *376*(26), 2523–2533. https://doi.org/10.1056/NEJMoa1612999

Chekroud, A. M. (2015). Unifying treatments for depression: an application of the Free Energy Principle. *Frontiers in Psychology*, *6*(FEB), 153. https://doi.org/10.3389/FPSYG.2015.00153/BIBTEX

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: a review of computational research. *Neuroscience & Biobehavioral Reviews*, *55*, 247–267. https://doi.org/10.1016/j.neubiorev.2015.05.005

Cittern, D., Nolte, T., Friston, K., & Edalat, A. (2018). Intrinsic and extrinsic motivators of attachment under active inference. *PLOS ONE*, *13*(4), e0193955. https://doi.org/10.1371/journal.pone.0193955

Constant, A., Hesp, C., Davey, C. G., Friston, K. J., & Badcock, P. B. (2021). Why depressed mood is adaptive: A numerical proof of principle for an evolutionary systems theory of depression. *Computational Psychiatry*, *5*(1), 60–80. https://doi.org/10.5334/cpsy.70

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., & Friston, K. (2019). Regimes of expectations: an active inference model of social conformity and human decision making. *Frontiers in Psychology*, *10*(MAR), 679. https://doi.org/10.3389/FPSYG.2019.00679/BIBTEX

Coyne, J. C. (1976). Toward an interactional description of depression. *Psychiatry*, *39*(1), 28–

40. https://doi.org/10.1080/00332747.1976.11023874

Cruwys, T., Haslam, S. A., Dingle, G. A., Haslam, C., & Jetten, J. (2014). Depression and social identity: an integrative review. *Personality and Social Psychology Review*, *18*(3), 215–238. https://doi.org/10.1177/1088868314523839

Cuijpers, P., Muñoz, R. F., Clarke, G. N., & Lewinsohn, P. M. (2009). Psychoeducational treatment and prevention of depression: the “coping with depression” course thirty years later. *Clinical Psychology Review*, *29*(5), 449–458. https://doi.org/10.1016/j.cpr.2009.04.005

D’Iuso, D. A., Dobson, K. S., Beaulieu, L., & Drapeau, M. (2018). Coping and interpersonal functioning in depression. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *50*(4), 248–255. https://doi.org/10.1037/cbs0000112

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: a synthesis. *Journal of Mathematical Psychology*, *99*, 102447. https://doi.org/10.1016/j.jmp.2020.102447

Da Costa, L., Parr, T., Sengupta, B., & Friston, K. (2021). Neural dynamics under active inference: Plausibility and efficiency of information processing. *Entropy*, *23*(4), 454. https://doi.org/10.3390/e23040454

Dobson, K. S., Quigley, L., & Dozois, D. J. A. (2014). Toward an integration of interpersonal risk models of depression and cognitive-behaviour therapy. *Australian Psychologist*, *49*(6), 328–336. https://doi.org/10.1111/ap.12079

Everaert, J., Podina, I. R., & Koster, E. H. W. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review*, *58*, 33–48. https://doi.org/10.1016/j.cpr.2017.09.005

Evraire, L. E., & Dozois, D. J. A. (2011). An integrative model of excessive reassurance seeking and negative feedback seeking in the development and maintenance of depression. *Clinical Psychology Review*, *31*(8), 1291–1303. https://doi.org/10.1016/j.cpr.2011.07.014

Feldman, A. G., & Levin, M. F. (1995). The origin and use of positional frames of reference in motor control. *Behavioral and Brain Sciences*, *18*(4), 723–744. https://doi.org/10.1017/S0140525X0004070X

Fernández-Theoduloz, G., Paz, V., Nicolaisen-Sobesky, E., Pérez, A., Buunk, A. P., Cabana, Á., & Gradin, V. B. (2019). Social avoidance in depression: a study using a social decision-making task. *Journal of Abnormal Psychology*, *128*(3), 234–244. https://doi.org/10.1037/abn0000415

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O☒Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, *68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. *Neural Computation*, *29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Gao, S., Assink, M., Cipriani, A., & Lin, K. (2017). Associations between rejection sensitivity and mental health outcomes: A meta-analytic review. *Clinical Psychology Review*, *57*, 59–74. https://doi.org/10.1016/j.cpr.2017.08.007

Gariépy, G., Honkaniemi, H., & Quesnel-Vallée, A. (2016). Social support and protection from depression: systematic review of current findings in Western countries. *British Journal of Psychiatry*, *209*(4), 284–293. https://doi.org/10.1192/bjp.bp.115.169094

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1803), 20190502. https://doi.org/10.1098/rstb.2019.0502

Gray, J. P., Müller, V. I., Eickhoff, S. B., & Fox, P. T. (2020). Multimodal abnormalities of brain structure and function in major depressive disorder: A meta-analysis of neuroimaging studies. *American Journal of Psychiatry*, *177*(5), 422–434. https://doi.org/10.1176/appi.ajp.2019.19050560

Hames, J. L., Hagan, C. R., & Joiner, T. E. (2013). Interpersonal processes in depression. *Annual Review of Clinical Psychology*, *9*, 355–377. https://doi.org/10.1146/annurev-clinpsy-050212-185553

Heins, C., Millidge, B., Demekas, D., Klein, B., Friston, K., Couzin, I. D., & Tschantz, A. (2022). pymdp: A Python library for active inference in discrete state spaces. *Journal of Open Source Software*, *7*(73), 4098. https://doi.org/10.21105/joss.04098

Huys, Q. J. M., Guitart-Masip, M., Dolan, R. J., & Dayan, P. (2015). Decision-theoretic psychiatry. *Clinical Psychological Science*, *3*(3), 400–421. https://doi.org/10.1177/2167702614562040

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, *112*(4), 323–343. https://doi.org/10.1007/s00422-018-0753-2

Kasch, K. L., Rottenberg, J., Arnow, B. A., & Gotlib, I. H. (2002). Behavioral activation and inhibition systems and the severity and course of depression. *Journal of Abnormal*

Psychology, 111(4), 589–597. https://doi.org/10.1037/0021-843X.111.4.589

Kirchner, L., Eckert, A., & Berg, M. (2022). From broken models to treatment selection: Active inference as a tool to guide clinical research and practice. *Clinical Psychology in Europe*, *4*(2), 0–4. https://doi.org/10.32872/cpe.9697

Kirchner, L., Schummer, S. E., Krug, H., Kube, T., & Rief, W. (2022). How social rejection expectations and depressive symptoms bi-directionally predict each other – A cross-lagged panel analysis. *Psychology and Psychotherapy: Theory, Research and Practice*, *95*(2), 477–492. https://doi.org/10.1111/papt.12383

Klerman, G. L., & Weissman, M. M. (1994). *Interpersonal psychotherapy of depression: a brief, focused, specific strategy*. https://rowman.com/ISBN/9781568213507/Interpersonal-Psychotherapy-of-Depression-A-Brief-Focused-Specific-Strategy

Kouvaris, K., Clune, J., Kounios, L., Brede, M., & Watson, R. A. (2015). How Evolution Learns to Generalise: Principles of under-fitting, over-fitting and induction in the evolution of developmental organisation. *Nature*, *388*, 539–547. https://doi.org/10.48550/arXiv.1508.06854

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: a predictive processing perspective. *Biological Psychiatry*, *87*(5), 388–398. https://doi.org/10.1016/j.biopsych.2019.07.017

Kupferberg, A., Bicks, L., & Hasler, G. (2016). Social functioning in major depressive disorder. *Neuroscience and Biobehavioral Reviews*, *69*, 313–332. https://doi.org/10.1016/j.neubiorev.2016.07.002

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, *27*(4), 986–1005. https://doi.org/10.1214/aoms/1177728069

Lipsitz, J. D., & Markowitz, J. C. (2013). Mechanisms of change in interpersonal therapy (IPT). *Clinical Psychology Review, 33*(8), 1134–1147. https://doi.org/10.1016/j.cpr.2013.09.002

MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation*, *4*(4), 590–604. https://doi.org/10.1162/neco.1992.4.4.590

Mccullough, J. P. (2006). *Treating chronic depression with disciplined personal involvement: cognitive behavioral analysis system of psychotherapy (CBASP)*. Springer.

McCullough, J. P. (2003). Treatment for chronic depression using cognitive behavioral analysis system of psychotherapy (CBASP). *Journal of Clinical Psychology*, *59*(8), 833–846. https://doi.org/10.1002/jclp.10176

McCullough, J. P. (2006). Healing interpersonal trauma using the interpersonal discrimination exercise. In *Treating Chronic Depression with Disciplined Personal Involvement* (pp. 123–159). Springer US. https://doi.org/10.1007/978-0-387-31066-4_6

McCullough, J., Schramm, E., & Penberthy, J. K. (2014). CBASP as a distinctive treatment for persistent depressive disorder. In *CBASP as a Distinctive Treatment for Persistent Depressive Disorder*. Routledge. https://doi.org/10.4324/9781315743196

Mikulincer, M., & Shaver, P. R. (2012). An attachment perspective on psychopathology. *World Psychiatry, 11*(1), 11–15. https://doi.org/10.1016/j.wpsyc.2012.01.003

Muñoz, R. F., Ghosh Ippen, C., Rao, S., Le, H.-N., & Valdes Dwyer, E. (2000). *Manual for group cognitive-behavioral therapy of major depression*.

Murphy, B., & Bates, G. W. (1997). Adult attachment style and vulnerability to depression. *Personality and Individual Differences*, *22*(6), 835–844. https://doi.org/10.1016/S0191-8869(96)00277-2

Nagy, E., & Moore, S. (2017). Social interventions: an effective approach to reduce adult depression? *Journal of Affective Disorders*, *218*, 131–152. https://doi.org/10.1016/j.jad.2017.04.043

Nanni, V., Uher, R., & Danese, A. (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *American Journal of Psychiatry*, *169*(2), 141–151. https://doi.org/10.1176/appi.ajp.2011.11020335

Nestor, B. A., Sutherland, S., & Garber, J. (2022). Theory of mind performance in depression: A meta-analysis. *Journal of Affective Disorders*, *303*, 233–244. https://doi.org/10.1016/j.jad.2022.02.028

Panitz, C., Endres, D., Buchholz, M., Khosrowtaj, Z., Sperl, M. F. J., Mueller, E. M., Schubö, A., Schütz, A. C., Teige-Mocigemba, S., & Pinquart, M. (2021). A revised framework for the investigation of expectation update versus maintenance in the context of expectation violations: The ViolEx 2.0 model. *Frontiers in Psychology*, *12*(November). https://doi.org/10.3389/fpsyg.2021.726432

Pettem, O., West, M., Mahoney, A., & Keller, A. (1993). Depression and attachment problems. *Journal of Psychiatry & Neuroscience : JPN*, *18*(2), 78–81. https://pubmed.ncbi.nlm.nih.gov/8461287/

Platt, B., Kadosh, K. C., & Lau, J. Y. F. (2013). The role of peer rejection in adolescent depression. *Depression and Anxiety*, *30*(9), 809–821. https://doi.org/10.1002/da.22120

Porcelli, S., Van Der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., Mogavero, F., Lobo, A., Olivera, F. J., Lobo, E., Posadas, M., Dukart, J., Kozak, R., Arce, E., Ikram, A., Vorstman, J., Bilderbeck, A., Saris, I., Kas, M. J., & Serretti, A. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews*, *97*, 10–33. https://doi.org/10.1016/j.neubiorev.2018.09.012

Rantala, M. J., Luoto, S., Krams, I., & Karlsson, H. (2018). Depression subtyping based on evolutionary psychiatry: Proximate mechanisms and ultimate functions. *Brain, Behavior, and Immunity*, *69*, 603–617. https://doi.org/10.1016/j.bbi.2017.10.012

Renner, F., Arntz, A., Leeuw, I., & Huibers, M. (2013). Treatment for chronic depression using schema therapy. *Clinical Psychology: Science and Practice*, *20*(2), 166–180. https://doi.org/10.1111/cpsp.12032

Rief, W., & Joormann, J. (2019). Revisiting the cognitive model of depression: the role of expectations. *Clinical Psychology in Europe*, *1*(1), 1–19. https://doi.org/10.32872/cpe.v1i1.32605

Roediger, E., Stevens, B. A., & Brockman, R. (2018). *Contextual schema therapy: An integrative approach to personality disorders, emotional dysregulation, and interpersonal functioning. - PsycNET*. New Harbinger Publications. https://psycnet.apa.org/record/2018-30793-000

Roepke, A. M., & Seligman, M. E. P. (2016). Depression and prospection. *British Journal of Clinical Psychology*, *55*(1), 23–48. https://doi.org/10.1111/bjc.12087

Safra, L., Chevallier, C., & Palminteri, S. (n.d.). *Social information impairs reward learning in depressive subjects: behavioral and computational characterization*. https://doi.org/10.1101/378281

Safra, L., Chevallier, C., & Palminteri, S. (2019). Depressive symptoms are associated with blunted reward learning in social contexts. *PLOS Computational Biology*, *15*(7), e1007224. https://doi.org/10.1371/journal.pcbi.1007224

Schramm, E., Klein, D. N., Elsaesser, M., Furukawa, T. A., & Domschke, K. (2020). Review of dysthymia and persistent depressive disorder: history, correlates, and clinical implications. *The Lancet Psychiatry*, *7*(9), 801–812. https://doi.org/10.1016/S2215-0366(20)30099-7

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., & Friston, K. (2015). Optimal inference with suboptimal models: Addiction and active Bayesian inference. *Medical Hypotheses*, *84*(2), 109–117. https://doi.org/10.1016/j.mehy.2014.12.007

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *ELife*, *8*, e41703. https://doi.org/10.7554/eLife.41703

Segrin, C. (1990). A meta-analytic review of social skill deficits in depression. *Communication Monographs*, *57*(4), 292–308. https://doi.org/10.1080/03637759009376204

Segrin, C. (2000). Social skills deficits associated with depression. *Clinical Psychology Review*, *20*(3), 379–403. https://doi.org/10.1016/S0272-7358(98)00104-4

Seidel, E.-M., Habel, U., Finkelmeyer, A., Schneider, F., Gur, R. C., & Derntl, B. (2010). Implicit and explicit behavioral tendencies in male and female depression. *Psychiatry Research*, *177*(1–2), 124–130. https://doi.org/10.1016/j.psychres.2010.02.001

Seligman, M. E. P. (1972). Learned helplessness. *Annual Review of Medicine*, *23*(1), 407–412. https://doi.org/10.1146/annurev.me.23.020172.002203

Sloman, L. (2003). Evolved mechanisms in depression: the role and interaction of attachment and social rank in depression. *Journal of Affective Disorders*, *74*(2), 107–121. https://doi.org/10.1016/S0165-0327(02)00116-7

Smith, R., Alkozei, A., Killgore, W. D. S., & Lane, R. D. (2018). Nested positive feedback loops in the maintenance of major depression: An integration and extension of previous models. *Brain, Behavior, and Immunity*, *67*, 374–397. https://doi.org/10.1016/J.BBI.2017.09.011

Smith, R., Badcock, P., & Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences*, *75*(1), 3–13. https://doi.org/10.1111/pcn.13138

Smith, R., Friston, K. J., & Whyte, C. J. (2022a). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, *107*, 102632. https://doi.org/10.1016/j.jmp.2021.102632

Smith, R., Friston, K. J., & Whyte, C. J. (2022b). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, *107*, 102632. https://doi.org/10.1016/j.jmp.2021.102632

Smith, R., Kirlic, N., Stewart, J. L., Touthang, J., Kuplicki, R., Khalsa, S. S., Feinstein, J., Paulus, M. P., & Aupperle, R. L. (2021). Greater decision uncertainty characterizes a transdiagnostic patient sample during approach-avoidance conflict: a computational modelling approach. *Journal of Psychiatry and Neuroscience*, *46*(1), E74–E87. https://doi.org/10.1503/jpn.200032

Smith, R., Schwartenbeck, P., Stewart, J. L., Kuplicki, R., Ekhtiari, H., & Paulus, M. P. (2020). Imprecise action selection in substance use disorder: Evidence for active learning impairments when solving the explore-exploit dilemma. *Drug and Alcohol Dependence*, *215*(April). https://doi.org/10.1016/j.drugalcdep.2020.108208

Spruit, A., Goos, L., Weenink, N., Rodenburg, R., Niemeyer, H., Stams, G. J., & Colonnesi, C. (2020). The relation between attachment and depression in children and adolescents: a multilevel meta-analysis. *Clinical Child and Family Psychology Review*, *23*(1), 54–69. https://doi.org/10.1007/s10567-019-00299-9

Starr, L. R., & Davila, J. (2008). Excessive reassurance seeking, depression, and interpersonal rejection: A meta-analytic review. *Journal of Abnormal Psychology*, *117*(4), 762–775. https://doi.org/10.1037/a0013866

Stellrecht, N. E., Jr., T. E. J., & Rudd, M. D. (2006). Responding to and treating negative interpersonal processes in suicidal depression. *Journal of Clinical Psychology*, *62*(9), 1129–1140. https://doi.org/10.1002/jclp.20298

Suffel, A., Nagels, A., Steines, M., Kircher, T., & Straube, B. (2020). Feeling addressed! The neural processing of social communicative cues in patients with major depression. *Human Brain Mapping*, *41*(13), 3541–3554. https://doi.org/10.1002/hbm.25027

Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical Psychology Review*, *31*(7), 1156–1168. https://doi.org/10.1016/j.cpr.2011.07.007

Tse, W. S., & Bond, A. J. (2004). The impact of depression on social skills. *Journal of Nervous & Mental Disease*, *192*(4), 260–268. https://doi.org/10.1097/01.nmd.0000120884.60002.2b

Uphoff, E., Ekers, D., Robertson, L., Dawson, S., Sanger, E., South, E., Samaan, Z., Richards, D., Meader, N., & Churchill, R. (2020). Behavioural activation therapy for depression in adults. *Cochrane Database of Systematic Reviews*, *2020*(7), CD013305. https://doi.org/10.1002/14651858.CD013305.PUB2/FULL

Wald, A. (1947). An essentially complete class of admissible decision functions. *The Annals of Mathematical Statistics*, *18*(4), 549–555. https://doi.org/10.1214/aoms/1177730345

Wang, Y., Yang, L. Q., Li, S., & Zhou, Y. (2015). Game theory paradigm: a new tool for investigating social dysfunction in major depressive disorders. *Frontiers in Psychiatry*, *6*(SEP), 128. https://doi.org/10.3389/FPSYT.2015.00128/BIBTEX

Watson, P. J., & Andrews, P. W. (2002). Toward a revised evolutionary adaptationist analysis of depression: the social navigation hypothesis. *Journal of Affective Disorders*, *72*(1), 1–14. https://doi.org/10.1016/S0165-0327(01)00459-1

Weightman, M. J., Air, T. M., & Baune, B. T. (2014). A review of the role of social cognition in major depressive disorder. *Frontiers in Psychiatry*, *5*(NOV), 179. https://doi.org/10.3389/FPSYT.2014.00179/BIBTEX

Wiese, W. (2017). Action is enabled by systematic misrepresentations. *Erkenntnis*, *82*(6), 1233–1252. https://doi.org/10.1007/s10670-016-9867-x

Wirth, J. H. (2016). Methods for Investigating Social Exclusion. In *Social Exclusion* (pp. 25–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-33033-4_2

Zimmer-Gembeck, M. J. (2016). Peer rejection, victimization, and relational self-system processes in adolescence: toward a transactional model of stress, coping, and developing sensitivities. *Child Development Perspectives*, *10*(2), 122–127. https://doi.org/10.1111/cdep.12174