Theory-Based Intervention for Transforming Harmful Behavior:

The Goal-Directed Predictive Processing Approach

Pieter Van Dessel[1], and Yannick Boddez[1]

[1]Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium; E-mail: Pieter.VanDessel@UGent.be;

[2]Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium; E-mail: Yannick.Boddez@UGent.be

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article as published in Review of General Psychology. The final article will be available, upon publication, via its DOI:

Van Dessel, P., & Boddez, Y. (in press). Theory-Based Intervention for Transforming Harmful Behavior: The Goal-Directed Predictive Processing Approach. Review of General Psychology. https://doi.org/10.1177/10892680251386723

Corresponding Author: Pieter Van Dessel, Department of Experimental-Clinical and Health Psychology, Ghent University, H. Dunantlaan 2, 9000 Ghent, Belgium.

Abstract

Human behavior can sometimes cause harm to the behaving individual, other organisms, or their environment. Psychological science plays a crucial role in understanding and addressing these harmful behaviors by providing direction to behavior change interventions. However, the field of psychology encompasses numerous theories, and translating these theories into effective practices can be challenging. In this paper, we introduce the Goal-Directed Predictive Processing (GDPP) framework, an integrative approach that combines key principles from predictive processing and goal-directed theories. The GDPP framework bridges recent theoretical advancements with practical application, offering a user-friendly model to assist practitioners across various domains. It emphasizes assessing inference chains that underlie harmful behaviors in risk situations, mapping out alternative inference chains in relation to expected surprise and self-concept beliefs, and designing targeted interventions to influence these predictive processes. We illustrate the application of this framework in the context of harmful alcohol consumption and depressive behavior.

Keywords: behavior change; goal-directed theory; goal-directed predictive processing; harmful behavior; predictive processing

Theory-Based Intervention for Transforming Harmful Behavior: The Goal-Directed Predictive Processing Approach

Behavior lies at the heart of society, driving the attainment of goals at individual, organizational, and societal levels, from fostering personal well-being to achieving financial success and addressing global challenges like climate change (Biglan, 2015; Thaler & Sunstein, 2008). However, there are many instances in which behavior causes harm (de Graaf & Wiertz, 2019; Marteau, 2018). We define behavior broadly as any type of response to a stimulus (see De Houwer & Hughes, 2020, for a discussion of the benefits of this definition) and harmful behavior as behavior that is likely to cause physical or mental damage, where harmfulness is a continuous quality of behavior (i.e., behavior can be harmful to some extent, in some sense). Both covert behavior (e.g., thoughts and feelings observable only to the individual) and overt behavior (e.g., physical actions observable to others) can be harmful by causing physical or mental harm to the individual engaging in the behavior or to other organisms, highlighting the multifaceted nature of harmful behaviors.

Practitioners across various fields often aim to mitigate harmful behaviors. While several methods are employed without recourse to scientific rationale, psychological science can furnish evidence-based practices by accumulating scientific insights into the determinants of behavior change (Albarracín et al., 2024; Hagger et al., 2020; Michie & West, 2013). First, behavior can be explained by referencing (relations between) environmental factors (De Houwer & Hughes, 2020; Skinner, 1953). For example, a boy’s behavior of crying (overt) and feeling anxious (covert) can be attributed to the boy being in the presence of a dog that barks loudly, an environmental event. Second, behavior can also be explained through reference to processes at the mental level (Gardner, 1987). For instance, the boy's crying might be elucidated by the boy making inferences that the dog may bite them. Finally, behavior can also be explained by biological factors such as neural activity within complex brain networks

(Cotman & McGough, 2014). Environmental, cognitive, and biological factors interact and gaining insights into their behavioral effects is essential for understanding and influencing behavior.

In this paper, we focus on insights about the mental processes underlying behavior. A growing body of literature has pointed to a practicality crisis in cognitive psychology, whereby theories often lack relevance, accessibility, or applicability to real-world problems (Berkman & Wilson, 2021). Although cognitive theories can offer valuable heuristic insights, they frequently lack the contextual sensitivity and implementation guidance that practitioners need. Moreover, because mental processes are not directly observable, cognitive theories must be evaluated on indirect grounds, such as their ability to clarify behavioral patterns (heuristic value) or to reliably predict and change behavior (predictive and influence value). As illustrated by the replication crisis, such evaluations must be based on findings that are replicable and generalizable (Open Science Collaboration, 2015).

To address these challenges, we propose a framework-based approach. Theoretical frameworks serve the purpose to organize and connect evidence-based assumptions into a coherent structure that guides practical application. The goal-directed predictive processing (GDPP) framework is such an effort. It draws from empirical findings with strong empirical backing and recent cognitive theories to provide structured guidance for understanding and influencing harmful behavior. By doing so, it aims to be both scientifically grounded and usable in real-world contexts.

For example, whereas therapeutic approaches such as Cognitive Behavioral Therapy (CBT) offer effective techniques for modifying harmful behavior, the GDPP framework provides a unified cognitive model that helps explain why these techniques work and how they can be adapted or extended. The framework combines well-supported assumptions about mental processes into a practical model that can inform interventions. In what follows, we first outline the foundations of the GDPP framework, then specify its core assumptions, connect it to robust empirical findings, and finally present practical directions for its implementation.

Goal-Directed and Predictive Processing Theories

At the mental process level, two hypotheses seem to stand out for having garnered significant support. First, behavior depends on inference, that is, the activation of beliefs based on their compatibility with other beliefs (Ajzen, 1991; Kube & Rozenkrantz, 2021). A belief is defined here as a mental representation that constitutes a statement about the world (also referred to as a ‘proposition’). Beliefs have truth value (i.e., one can evaluate them as being true or false) but are not necessarily verbal in nature (e.g., they can involve embodied representations: De Houwer, 2014). Second, mental representations of desired outcomes, or goals, are fundamental determinants of behavior (Locke & Latham, 2006; Moors et al., 2017). While these concepts offer a solid foundation for practice, on their own they may be too broad or imprecise. Fortunately, they are integral components of several established theories, most notably predictive processing and goal-directed theories, which incorporate these ideas at their core and have garnered substantial support in recent years.

Predictive Processing Theories

The predictive processing framework is a broad theoretical approach that has had many different implementations over time (Sprevak & Smith, 2023). It revolves around the basic idea that behavior is driven by predictions (i.e., inferences about future events such as sensory input) and that these predictions are continually updated based on prediction error, that is, the disparity between expected and actual input (Friston, 2010). Originally influential in explaining perception (Helmholtz, 1962), predictive processing has more recently expanded to explain other behaviors under the framework of active inference (Clark, 2016; Parr et al., 2022). Here, behavior results from the process of aligning the environment with predictions (e.g., if we predict our home to be well-organized, we act to make it so). This process is guided by the overarching principle of entropy reduction (Friston, 2010), a foundational principle that is also observed in physical systems. It refers to the mental system's inherent tendency to reduce the entropy or uncertainty within the (belief-based) system - where mental system denotes a theoretical construct responsible for generating and updating mental representations such as beliefs (see Table 1 for applied definitions). In the context of harmful behavior, this means that individuals engage behavior such as substance use or rumination, when these are predicted to best reduce uncertainty in familiar contexts.

Table 1.

Applied definitions used in this paper

Term	Applied definition
Belief updating	The process of revising activated beliefs to reduce expected surprise, based on comparisons between alternative belief candidates.
Entropy	The average uncertainty within a belief system; higher entropy reflects greater disorder in the model.
Expected surprise	The anticipated level of surprise that would result from future inputs, given a specific belief update.
Mental system	A theoretical construct referring to the functional system that generates and updates mental representations (e.g., beliefs). It is situated at the cognitive level and should not be conflated with the biological brain, which operates at a different level of explanation.
Precision	The confidence assigned to a prediction; determines how much weight is given to a prediction versus the incoming sensory input.
Prediction error	The mismatch between what is expected and what is sensed at a given moment.
Surprise	The degree to which a sensory input is unexpected, given current beliefs.

Note. Definitions are provided in applied terms to support intervention design; more formal treatments are available in Friston (2010) and Parr et al. (2022).

Predictive processing theories are firmly rooted in behavioral and neural evidence, demonstrating applicability across various neural and behavioral responses (Clark, 2016; Hohwy, 2013). However, despite offering a comprehensive understanding of behavior at both neural and cognitive levels, these theories are often described in complex computational terms, which can hinder practical application. While they make specific predictions at the neural level and can be insightful for explaining laboratory-based actions, most implications may not be suitable for making specific behavioral predictions or for application to real-life behavior (Friston et al., 2016; Sprevak & Wilkinson, 2021; see also Clark, 2024). Moreover, in practical contexts, the beliefs that drive behavior seem strongly person- and context-dependent (Hayes et al., 2012; De Houwer & Hughes, 2020), making computation-based and neural-level predictions difficult to achieve. These limitations constrain the practical utility of traditional predictive processing theories in designing behavior change interventions. While goals could potentially serve as entry points for theory-based interventions, their role is sometimes unclear in predictive processing theories, with some implementations recasting goals as precise predictions (Clark, 2020; Pezzulo et al., 2018).

Goal-Directed Theories

Goal-directed theories propose that behavior is driven by the activation of mental representations of behavioral outcomes and their value (Kruglanski & Szumowska, 2020; Moors et al., 2017). A recent theory by Agnes Moors (2022) explains action through a goal-directed cycle that begins with comparing a representation of the current state of affairs (e.g., the state of our house) with a representation of a desired outcome (i.e., a goal; e.g., a clean house). Detecting a discrepancy may then lead to (a) reinterpretation of the current state (immunization; e.g., perceiving the house as sufficiently clean), (b) choosing a different goal (accommodation; e.g., deciding that having a clean house is not important), or (c) selecting an action based on its expected utility (assimilation; e.g., cleaning the house), with the latter potentially resulting in the observable action. This cycle also helps explain harmful actions: for instance, someone experiencing stress may compare their current state with a goal of relief, and infer that drinking alcohol will achieve it, even if the longer-term consequences are damaging.

Goal-directed theories provide a robust framework for understanding and predicting behavior and have been applied across diverse domains, including addiction (Buabang et al., 2024), open-access publishing (Köster et al., 2021), sustainable consumption (Vermeir et al., 2020), and weak-willed behavior (Moors, 2019). These theories are sufficiently broad to encompass various (overt) behavioral phenomena while remaining operationalizable for practical applications. Unlike broader cognitive theories, such as those emphasizing general mental constructs or the general idea that beliefs and goals underlie behavior (e.g., Ajzen, 1991; Kahneman, 2011), which are often abstract and challenging to operationalize (Albarracín et al., 2024), goal-directed theories enable the delineation of context-dependent process chains. This makes them well-suited for developing actionable and targeted interventions that translate theoretical insights into practical strategies.

However, the practical utility of current implementations of goal-directed theories may be constrained by the absence of a general driving force, such as entropy reduction. This driving force holds practical value as it can serve as a guiding principle for mapping situation-specific inferential processes and facilitating the development of tangible interventions for complex behaviors. Relatedly, goal-directed theories place less emphasis on the fundamental process of prediction (Munakata et al., 2023) compared to predictive processing theories. They therefore seem to focus less on self-concept beliefs, which are considered to play a crucial role in shaping behavior (Bandura, 1977; Rosenberg, 1979). It might therefore have value to build a predictive processing framework that incorporates elements of goal-directed frameworks to enhance explanatory power and practical applicability.

Integrating Predictive Processing and Goal-Directed Theories

Given the strengths and limitations of both types of theories, we propose a framework that elucidates key assumptions of predictive processing theories about mental (rather than neural) processes while integrating goal-directed action. Our aim is not to strictly adhere to all the specific assumptions of predictive processing or goal-directed theories. Instead, we seek to extract and integrate those aspects that offer high heuristic value based on a robust evidence base and are most practical for understanding, predicting, and influencing real-life behavior. By merging their strengths and highlighting how the two types of theories can align, particularly in emphasizing the role of inferences and desired outcomes in driving behavior, we create a user-friendly framework with a foundation in recent empirical evidence. This approach bridges the gap between complex theoretical models and practical applications, making it more accessible for practitioners to implement theory-based interventions. For practitioners aiming to address harmful behaviors, this integration highlights not only why such behaviors persist but also which inferential steps can be targeted to promote adaptive change.

The Goal-Directed Predictive Processing Framework

Core Assumptions

Mental model of the environment and prior beliefs. In any given situation, individuals maintain a mental model of their internal and external environment, composed of prior beliefs about themselves and the world around them (Clark, 2016). For example, when entering a bar, a person who struggles with alcohol use may activate a mental model that includes beliefs such as “in this place I usually drink,” “my friends will probably be drinking too,” and “I feel relaxed when I have a beer in this place.” These beliefs represent key features of the situation, including both external circumstances and internal tendencies, that shape how the individual interprets what is happening.

Predictions and sensory signals. From this mental model, individuals generate predictions about the internal signals they will encounter, that is, incoming sensory and visceral inputs. These predictions are assigned a level of precision (i.e., confidence in the prediction), allowing for some margin of variability or 'noise' in the signal. When actual input deviates from these predictions beyond the tolerated margin, a 'prediction error' is registered. For instance, if the person enters the bar and does not experience the predicted sensations, such as the feel of holding a drink, the taste of alcohol, or the relaxing bodily state, the discrepancy between prediction and input gives rise to prediction error.

Minimizing expected surprise and belief updating. When significant prediction error is registered, it signals that the current mental model is no longer sufficient to account for the environment. The system then faces increased expected surprise: it anticipates greater uncertainty about future inputs unless the model is adjusted. This uncertainty is disadvantageous because it reflects a disorganized model that demands more cognitive and metabolic resources to generate accurate predictions (Friston, 2010). In response, the mental system evaluates potential updates to the beliefs that constitute the model (e.g., based on Bayesian reasoning principles: Gershman et al., 2015; Sanborn & Chater, 2016) and selects those that are most likely to reduce future prediction errors, that is, the updates that minimize expected surprise (Parr et al., 2022). For example, in the bar setting, if the person does not experience the expected sensations of drinking or feeling relaxed, they may revise their beliefs to better align with the current input. They might consider that the bar is unusually quiet, that they are more tense than usual, or that they will drink alcohol, and evaluate which belief update best reduces expected surprise.

Active inference and behavior. In some cases, instead of updating beliefs about the external world, the system resolves prediction error by updating beliefs about the system itself, such as about its behavior. For instance, an individual may generate a prediction that they will perform a specific action, which sets in motion lower-level predictions of motor states that directly trigger an action, a process known as active inference (Friston, 2010). For example, in the bar setting, the person may infer that the belief "I will drink alcohol" minimizes surprise, which leads to predictions about specific actions (e.g., ordering a beer) that initiate the motor behavior of drinking.

Self-concept beliefs and goals. In the current framework, goals shape the inferential structure of the entire system. Every belief that is activated is evaluated with reference to expected surprise, which itself is qualified by current goals. That is, a "wanted" outcome is inherently one that is expected to reduce surprise. Thus, goals are conceptualized as beliefs about outcomes that are valuable with respect to entropy minimization and that should therefore be pursued. As a result, goals determine which beliefs are activated, which predictions are made, and how much precision is assigned to predictions about specific input.

Because of this, the beliefs a person holds about themselves, their self-concept beliefs, play a central role in shaping behavior. These include assumptions about who they are, what they are capable of, and what outcomes they strive for. For example, in the bar scenario, whether the person ultimately infers that they will drink depends on their broader self-concept: if they hold the belief "I am someone who drinks to relax," then drinking aligns with their goal of achieving relaxation and is seen as a coherent next action. In contrast, if they believe "I'm trying to avoid alcohol," then the same prediction error might activate different belief updates or behaviors. Thus, behavior reflects an inference about what the self will do next, given its goals and identity. In this sense, the framework integrates the goal-directed tradition's emphasis on wanted outcomes with predictive processing's emphasis on prediction and entropy reduction.

The Mental Process Underlying Behavior

From the goal-directed predictive processing (GDPP) perspective, the mental process that drives behavior can be described as a five-step process (Figure 1).

1. *Model of the environment*. Individuals form a mental model of their internal and external environment. This model is constructed on the basis of (goal-related) beliefs that are activated in the current context and fosters predictions about internal signals with a certain level of precision.

2. *Surprise.* Predictions about internal signals are compared to the internal signals. These comparisons generate a certain level of surprise or prediction error.
3. *Estimation of expected surprise reduction.* In response, individuals evaluate the beliefs that form the foundation for their mental model (e.g., belief A about the environment) and consider alternative beliefs (e.g., belief B about engaging in specific behavior). They assess how updating these beliefs would influence the prediction errors they expect to occur in the future ("*expected surprise*").
4. *Comparison of expected surprise reduction*. The reduction of expected surprise is compared for the potential updates in beliefs (e.g., for updating belief A and B).
5. *Taking action.* If updating a particular belief about engaging in specific behavior (e.g., belief B) is predicted to minimize expected surprise more effectively than other options (e.g., updating belief A), the individual implements this belief update, resulting in the corresponding behavior.

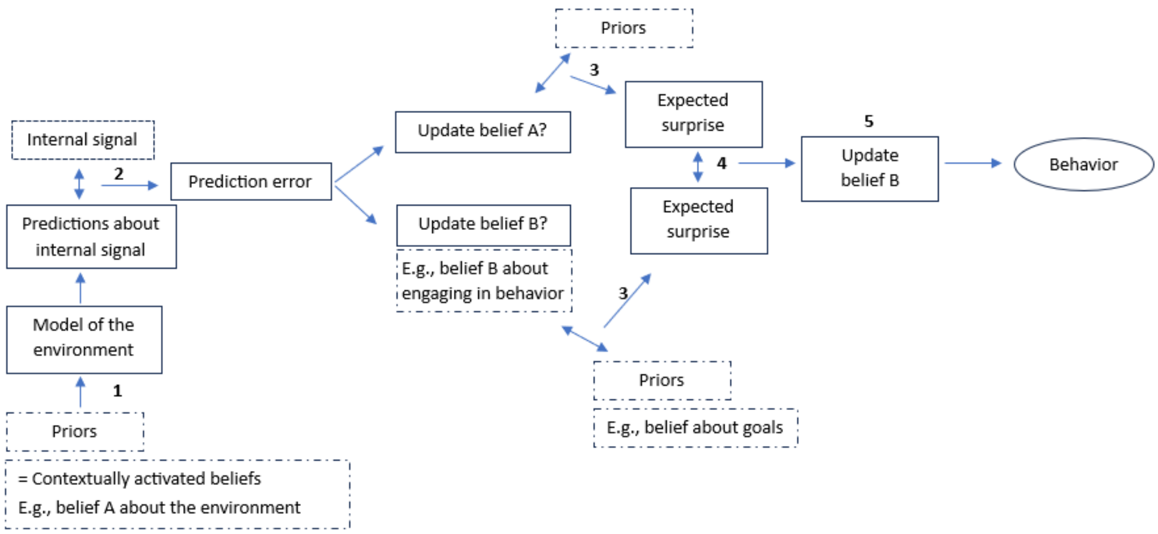

Figure 1. In the GDPP model, behavior results from a five-step process. First, people build a mental model of the environment based on contextually activated prior beliefs. Second, prediction error is registered following the comparison of predictions about internal signals and actual internal signals. In the third and fourth steps, expected surprise is calculated and compared across candidate belief updates. Finally, the belief that is expected to minimize surprise most effectively is updated and, when this constitutes a belief about engaging in behavior, the corresponding behavior is emitted.

Heuristic Power

While the GDPP framework does not rely on a large number of complex assumptions, it offers significant explanatory value for key empirical findings in psychological science. Below, we discuss some of the most robust and widely replicated findings that are particularly relevant for practitioners and thus essential for any theoretical framework to capture.

Belief bias. An abundance of research demonstrates that behavior is systematically shaped by beliefs, manifesting as various biases such as confirmation, interpretation, and perception bias (Gilovich et al., 2002; Kube & Rozenkrantz, 2021). Within the GDPP framework, such biases arise because prior beliefs constitute the mental model of the world from which behavioral predictions are derived. Since behavior serves to minimize expected surprise, perception and interpretation (covert behaviors) are biased toward confirming deeply ingrained, frequently activated beliefs (i.e., hierarchically high beliefs; Pezzulo et al., 2018). Even when faced with significant prediction errors, individuals may avoid updating high-level beliefs, as doing so would increase expected surprise. Instead, they engage in behaviors that confirm existing beliefs (e.g., confirmation bias; Chinn & Brewer, 1993), reinforcing the model and preserving cognitive equilibrium.

Goal-directed behavior. A wealth of evidence underscores the pivotal role of action outcomes in shaping behavior (Thorndike, 1905; De Houwer et al., 2024) and shows that people often engage in actions that yield desired outcomes, aligning with their self-reported goals (Locke & Latham, 1990). The GDPP framework explains this pattern by emphasizing the role of self-concept beliefs, beliefs about the kind of person one is and the behaviors one typically engages in. These beliefs guide predictions about which actions are most consistent with one’s goals and identity, and thus most likely to minimize expected surprise. This aligns with broader evidence showing that self-concept beliefs play a central role in behavior regulation (Bem, 1972; Deci & Ryan, 2012).

An important aspect of the GDPP model is that high-level self-concept beliefs, beliefs that are easily activated and tightly interconnected with frequently accessed beliefs (akin to hyperpriors; Clark, 2013), exert a strong influence on behavior. These beliefs may have evolutionary origins (e.g., the expectation that safety is a valid goal), but can also emerge through personal learning. For instance, a child consistently given blue cutlery in daycare may develop the belief that blue items “belong to me.” This self-concept belief can drive future behavior, such as requesting blue utensils, to fulfill the predicted outcome of possessing blue items. As such beliefs become more generative, that is, more often activated to minimize expected surprise, they may be represented at higher levels in the belief hierarchy and more strongly connected to other beliefs. Updating them would then disrupt the internal coherence of the system and increase expected surprise, making them more resistant to change.

Emotions. Another foundational insight is that humans experience a wide range of emotions, which significantly influence behavior (Baumeister et al., 2007). Within the GDPP framework, emotions are conceptualized as a type of covert behavior. Like other behaviors, emotions arise from predictions of their occurrence. The subjective component of emotion, the conscious experience of a particular emotional state, is conceptualized as the activation of a meta-belief: the belief that one is experiencing a given emotional state (Cleeremans, 2011).

The valence-related quality of emotions can be understood in terms of prediction error. Negative emotions are linked to mental distress, a state in which there is high expected surprise without a clear path for resolution that does not perturb the mental system (Carver & Scheier, 1990; Van De Cruys, 2017; Van De Cruys & Van Dessel, 2021). For instance, a person may feel anxious because they predict experiencing this emotion when anticipating an uncontrollable event that increases expected surprise. Positive emotions, in contrast, are associated with states that reduce expected surprise. Importantly, a surprising event is not inherently negative. Its emotional valence depends on how the system predicts the event will affect future surprise. A surprising event may feel positive if the individual predicts they can cope with it effectively and that it will ultimately reduce expected surprise over time.

Emotions are thus intricately connected to desired states and deeply rooted in individuals' goals (Bagozzi & Pieters, 1998; Moors, 2022; Moors et al., 2017). As such, they also play a significant role in priming other behaviors, such as thoughts and actions. Specifically, emotions may prompt the prediction that engaging in behaviors that promote or alleviate these emotions minimizes expected surprise. For example, someone experiencing distress may predict that drinking alcohol or avoiding a confrontation will reduce that distress, thereby reinforcing this behavior (Hogarth, 2020; Lovibond, 2006).

Cognitive dissonance. A related empirical observation is that people often experience discomfort due to inconsistencies in their beliefs and/or behaviors, a phenomenon known as cognitive dissonance (Festinger, 1957). This discomfort typically motivates individuals to modify their beliefs or behaviors to resolve the inconsistency (Harmon-Jones & Harmon-Jones, 2007; Moors et al., 2017). Within the GDPP framework, this discomfort arises because the system predicts that holding two conflicting beliefs or engaging in contradictory actions increases expected surprise. To minimize expected surprise, the system may update one of the conflicting elements. For instance, someone who values sustainability but uses single-use plastics may either change their behavior or adjust their beliefs about the impact of their actions to reduce the inconsistency and its related emotion (Moors et al., 2017).

Harmful behavior. Individuals often engage in behavior that appears harmful, and such behavior is frequently attributed to irrationality, mental dysfunction, or non-goal-directed processes (Leshner, 1997; Kahneman & Tversky, 1979). In contrast, the GDPP framework posits that all behavior, including harmful behavior, serves a function within the person's active mental model. Specifically, the behavior is predicted to minimize expected surprise relative to the current belief structure, even if that behavior conflicts with long-term or self-reported goals.

For instance, a person who self-harms may hold the belief that this behavior is their way of coping with feelings of stress. When these feelings are experienced, the system may predict that self-harm is the most effective way to reduce mental distress, even if the person explicitly endorses a broader goal of physical well-being, for example in a therapeutic context. The harmful behavior thus reflects the immediate resolution of prediction error based on available beliefs, not a breakdown in functioning. This perspective highlights the importance of understanding which beliefs are active, how they are structured, and what goals they are tied to in order to design effective interventions, as discussed in the next section.

Implications for Behavior Change Practice

Adopting the GDPP model offers a novel approach for designing interventions aimed at reducing harmful behavior. Below, we outline four general principles rooted in GDPP theory, followed by specific intervention strategies and two applied examples related to harmful alcohol consumption and depressive behavior.

General Principles

1. Target the context-dependent automatic application of beliefs. In the GDPP framework, behavior results from the automatic application of beliefs to present contexts, shaped by the minimization of expected surprise: people act in ways that align with what they predict they will do, given their beliefs and the situation. The automaticity of this process refers to this process being characterized by features such as speed and efficiency – features that exist on a continuum such that automaticity is not an all-or-nothing phenomenon (Moors, 2016). Note that, within this framework, the (automatic) inferences that underlie behavior are distinct from conscious thoughts in the sense that they generate predictions about behavior, which can manifest as both overt actions and verbal behaviors, including conscious thoughts. Importantly, given the crucial role of the inferential process, behavior change interventions should not just target the validity of beliefs, but their situational activation, and more specifically, whether a person predicts they will act on those beliefs in relevant contexts.

This focus on belief activation marks a key distinction between GDPP and classical cognitive approaches, such as early forms of Cognitive Behavioral Therapy (CBT), which tend to view beliefs as relatively stable, conscious representations like schemas (Beck, 2011). Instead, GDPP emphasizes the inferential process through which beliefs are automatically applied in context: Will the belief be applied when the person is watching TV, in a social setting, or under stress? For instance, it may be less effective to convince someone that "eating healthy is good" than to promote the automatic inference that they will act on this belief when snacking during a stressful evening. Such interventions aim to shift belief-based predictions in context, not just belief content.

While CBT encompasses many effective techniques that often target maladaptive beliefs (David et al., 2018), and is therefore broadly compatible with the GDPP framework, it remains a therapeutic framework rather than a unified cognitive model. GDPP offers a unified account of why these techniques can work: they may reduce distress and maladaptive behavior by altering the prediction that harmful beliefs will guide action, and by strengthening goal-consistent alternatives. This perspective also aligns with findings linking CBT success to shifts in patient expectations (Kirsch, 1990), grounding such effects in a broader theory of inference and prediction. GDPP may thus help bridge the gap between theory and practice, offering CBT practitioners a principled, inference-based account of behavior change that highlights new approaches. For instance , rather than merely challenging the validity of maladaptive beliefs (or targeting related conscious thoughts), GDPP-guided interventions should aim to disrupt the automatic prediction that such beliefs will be applied in key contexts, and promote the inference that alternative beliefs will be applied instead.

2. Target inferences in reference to expected surprise. In the GDPP framework, people act in ways that minimize expected surprise. Interventions can thus promote behavior change by altering what is expected to be surprising. This can be achieved in two ways: by increasing expected surprise for harmful behaviors or by decreasing expected surprise for healthier alternatives.

The first approach involves targeting inferences that make maladaptive behavior appear less self-congruent or effective. For example, if a person expects that self-harm will reduce distress, an intervention might prompt new inferences, such as the idea that self-harm is inconsistent with their usual coping methods, thus increasing the expected surprise of engaging in that behavior. As the prediction becomes less stable, the behavior becomes less compelling.

The second strategy aims to reduce expected surprise for helpful alternatives. This involves strengthening inferences that make adaptive behaviors seem self-relevant. For instance, the person may come to expect that using a relaxation technique will reduce distress in a predictable and coherent way. As a result, the alternative behavior becomes more likely because it is inferred to align with the person's goals and expectations.

Together, these strategies shift behavior by reshaping the inferential landscape: harmful behavior becomes less expected, and adaptive behavior becomes a more coherent prediction. This highlights how GDPP-guided interventions work by modifying belief-based predictions about which behaviors will reduce expected surprise in a given situation.

3. Target goal-directed inferences in reference to self-concept beliefs. In GDPP, behavior results not from mere representations of wanted outcomes but from beliefs about one's goals and the actions one is likely to engage in. While traditional goal-directed models emphasize the role of outcome value and expectancy (Moors et al., 2017), GDPP highlights the importance of predicting that one will act in line with a goal, which may occur when that prediction is seen as most coherent with one's self-concept.

Interventions should therefore aim to shift the inference that certain behaviors are likely to occur because they align with the person's activated goals and self-related beliefs. For example, rather than solely emphasizing that social engagement can lead to better relationships, an intervention might prompt the inference: "Because I value connection and often act on that goal, I predict I'll reach out to a friend when I feel lonely." This framing draws on identity and consistency, making the behavior feel expected and self-relevant in the moment.

4. Target the different steps in the behavior process chain. To effectively change behavior, interventions must target the specific inferences that sustain it. In the GDPP framework, behavior emerges from a chain of context-sensitive inferences, beginning with predictions about internal states and extending to predictions about which actions will be performed to minimize expected surprise. Each step in this process chain offers a potential intervention point.

The first step involves comparing internal signals with predictions about those signals. This comparison generates surprise, which can be reduced either by modifying the signal itself or by adjusting the prediction. For example, in a prison context, an inmate might experience rising physiological arousal in an overcrowded cell and predict that this signals a violation of personal space. This mismatch creates prediction error that may drive predictions of harmful behavior such as aggression. Interventions could target the signal (e.g., reduce overcrowding) or promote alternative inferences (e.g., help the person predict that they are able to cope well with the limited space).

The next steps involve considerations of how to reduce expected surprise. A person may predict that aggression will resolve the tension and restore control, making it the expected behavior. To shift behavior, interventions can promote new inferences about alternative actions, such as taking a walk or using breathing techniques, as more effective and better aligned with the person's self-concept (e.g., "I am someone who deals with stress by walking to calm myself").

Designing such interventions requires first mapping the individual's current inference chain: What internal signals are registered? What predictions do they make about those signals? What (self-concept) beliefs are activated and what behaviors are inferred as the solution? Once this chain is identified, interventions can introduce new inferences that are both plausible within the person's belief system and capable of shifting behavior in context.

Intervention Strategies

GDPP-guided behavior change interventions focus on reshaping the predictions people make about their own future actions. These predictions emerge from context-sensitive inference chains, which connect internal states, beliefs, and anticipated behavior. Effective interventions must therefore assess and reconfigure these inference chains using a structured three-step process: (1) assessment, (2) mapping adaptive inference chains, and (3) intervention.

Step 1: Assessment. In line with therapeutic approaches such as CBT, the first phase involves a detailed analysis of the target behavior and its risk situations. However, from a GDPP perspective, assessment must go further: it should uncover the inferences and self-related beliefs individuals automatically apply in these contexts, which sustain the behavior by making it predictable and self-congruent.

One effective method for this is Applied Behavior Analysis (ABA), which tracks the relationship between environmental cues, behaviors, and consequences (Fischer et al., 2021). While ABA is often used in real-time settings, GDPP extends it by focusing on inferred beliefs (see also Vanaken et al., 2021). Retrospective discussions and verbal reports can help identify which beliefs are likely activated in specific situations although it should be noted that verbal expressions are not direct introspections but observable behaviors that also result from inferences and can inform hypotheses about the belief network in that sense.

The goal is to identify both (1) beliefs that maintain harmful behavior and (2) alternative beliefs expressed elsewhere (e.g., in other contexts or verbal reports) that are not applied in risk situations. These unused but accessible beliefs may form the building blocks for new, adaptive inference chains that sufficiently align with the individual's self-concept.

Step 2: Mapping adaptive inference chains. This step involves designing new inference chains that can override the harmful behavior chain by reducing expected surprise more effectively. These chains should integrate self-related beliefs identified in Step 1 and align with the individual's mental model, particularly their goals, identity, and sense of coherence. Interventions often fail because they bypass this step and rely on surface-level assumptions about behavior (Mertens et al., 2022; Milkman et al., 2022).

By contrast, GDPP-guided interventions require identifying each inferential step: What signal is interpreted? What belief is activated? What behavior is predicted? Which belief can be introduced at what point in the chain to shift the prediction? For example, if an individual experiences rising arousal and predicts that aggression is a coherent response, the new chain must introduce the belief: “When I feel tension, I predict I’ll go for a walk to calm myself.” The alternative action is more effective only if it is inferred to align with self-concept, context, and prior expectations.

Step 3: Intervention. The final step is to implement strategies that promote the new inference chains. Two complementary approaches can be distinguished based on their target: inference nudging and inference training.

A. Inference nudging. This involves modifying the environment where harmful behavior occurs to promote targeted inferences in that context. Like traditional nudging (Thaler & Sunstein, 2008), it involves modifying a choice environment, but with a focus on the beliefs and predictions activated in that moment (Van Dessel et al., 2022). For example, stimulus control interventions (Pierce & Cheney, 2018) may be effective only if they manage environmental cues that promote the application of relevant beliefs within the targeted inference chain in reference to expected surprise.

B. Inference training. This strategy involves modifying an environment to prepare individuals to apply adaptive inferences *beyond the immediate situation,* thereby "training" these inferences for future reference. Because maladaptive beliefs are often highly generative beliefs, new beliefs must be scaffolded to create a bridge between current and future beliefs, targeting the individual's "zone of proximal development" (Vygotsky, 1978).

Effective training fosters the prediction that the new behavior will occur in relevant future situations, which can be achieved through various techniques. One such technique is contingency management, which involves structuring the environment so that, within a given situation, certain behavior consistently leads to specific consequences (Fischer et al., 2021; Petry, 2011). For instance, a person's act to refrain from substance use may consistently be followed by access to a desired activity. Although this approach typically focuses on promoting observable actions, it can also be extended to managing contingencies related to thoughts and feelings, or the application of specific beliefs more generally. Moreover, while contingency management may promote inference chains about the presence of contingencies, it will be effective only to the extent that it fosters the prediction of relevant behavior change as the result of these contingencies.

Similarly, compatibility-focused techniques that provide arguments, such as through verbal discussion or experience, can support inference beliefs in the adaptive behavior inference chain or contrast with beliefs in the harmful behavior inference chain. This is common practice in therapeutic approaches such as CBT (David et al., 2018), acceptance and commitment therapy (Hayes et al., 2006) and exposure therapy (Abramowitz et al., 2019). In the GDPP framework, however, lasting behavior change should occur only when individuals infer that they will act (and are able to act) on the new beliefs, rather than when they merely register the information that the behavior contrasts with certain beliefs. Therefore, especially when addressing recurrent or habitual behavior, it is crucial that individuals draw relevant inferences about how this information will impact their own behavior (see Vanderschoot, & Van Dessel, 2022, for an illustration in the context of Post-Traumatic Stress Disorder).

Finally, there is potential in practice-based interventions that involve applying targeted beliefs in relevant situations, akin to cognitive bias modification (CBM) interventions that aim to change habitual behavior by targeting (bias in) automatic cognitive processes (Wiers et al., 2013). From the GDPP perspective, traditional CBM interventions, such as repeatedly practicing specific behaviors like avoiding alcoholic stimuli (Wiers et al., 2010), might sometimes activate relevant beliefs and promote action predictions. However, while the repetitive nature of these tasks promotes the sampling of these beliefs (similar to learning a skill like driving or doing math), this sampling may often fail to transfer to real-life behavior (Wiers et al., 2020). More effective practice-based intervention requires mapping out relevant inference chains and targeting the inference that one will later enact the practice in relevant situations given the better fit with a person's mental model (e.g., with beliefs about their goals: Van Dessel et al., 2018). To support this, the training environment may mirror real-life scenarios to establish cues that activate the learned inferences in relevant risk situations and information can be provided that supports the transfer of practiced inferences to real-world behavior (e.g., information about training efficacy) (Van Dessel et al., 2024).

Settings of Particular Relevance for Application

Because the GDPP framework models behavior as the outcome of predictions in light of goals, beliefs, and expected surprise, it can in principle be applied to any behavior. However, it may be particularly valuable in contexts where behavior change is difficult to achieve through standard approaches alone. This includes situations characterized by deeply ingrained habits, motivational ambivalence, or conflicting goal structures, as well as contexts where individuals want to change overt or covert responses (e.g., emotions, ruminative thoughts) but feel unable to do so.

Such contexts occur both at the individual and the population level. At the individual level, prominent examples include addiction, and impulsive behavior more generally, such as in the context of criminal justice (Linthout et al., 2025). In such cases, behavior change interventions may benefit most from an integrative approach that considers the automatic belief-based processes that prevent change. The GDPP framework can guide such interventions by offering a principled method to identify which beliefs should be targeted, and how belief updating can be facilitated by aligning interventions with a person's broader goal structure.

At the population level, similar logic applies. Examples include promoting sustainable behavior, such as in educational contexts, by identifying and targeting widely shared inference chains within the target group (Van Dessel et al., 2025). These shared inferences are often shaped by recurring contingencies in the environment, yet interventions rarely examine or target them directly. In the context of mental health promotion among adolescents, for instance, inference chains can be mapped by leveraging research on beliefs linked to depression symptoms (Beck & Dozios, 2011) and commonly held self-concept beliefs in the target population (e.g., Deci & Ryan, 2012). Interventions might then reinforce the contextual application of core beliefs, such as "every individual holds value." By promoting predictions consistent with such beliefs, these interventions may counteract depressive symptoms and encourage more adaptive behavior.

To illustrate these ideas in more depth, the next sections present two applied cases in which behavior change is particularly challenging and clinically relevant: harmful alcohol consumption and depressive behavior. These cases serve as detailed demonstrations of how

GDPP can be used to map maladaptive inference chains and design interventions that reconfigure them toward adaptive outcomes.

Applied Case: Harmful Alcohol Consumption

Alcohol addiction, like any other mental health condition, can be seen as a compilation of distinct patterns in overt and covert behavior that dynamically interact (Roefs et al., 2022). Focusing on the behavior of harmful alcohol consumption, the GDPP perspective outlines a five-step process:

1. The person constructs a mental model of their internal and external environment based on activated beliefs. For instance, a person may register sitting on the couch and feeling stressed.
2. Surprise is registered when comparing internal signals and predictions about these signals. For instance, if the person is at home feeling stressed without drinking a beer, this might conflict with their prediction that they typically drink beer in this setting and register internal sensations related to alcohol consumption.
3. The mental system evaluates which beliefs can be updated to reduce expected surprise. When the belief 'I will drink alcohol' is considered, the system infers its outcomes, which may heighten expected surprise and evoke sensations such as craving.
4. If no other beliefs are found that align with their beliefs and can be updated to better resolve surprise, updating of the belief to consume alcohol is estimated to best minimize surprise.
5. This inference triggers low-level motor predictions that lead to the behavior of alcohol consumption.

This chain highlights how harmful behavior is maintained by automatic, context-sensitive predictions. To intervene, GDPP proposes a three-step process that systematically reconfigures these inferences.

The first step involves identifying the key inference chains that sustain alcohol use in risky situations. For instance: “Stress is overwhelming in this situation” → “Alcohol reduces stress” → “I am likely to drink to reduce stress” → “I will drink in this situation.” Methods such as Applied Behavior Analysis (ABA) and verbal exploration of prior behavior can uncover these patterns, including alternative beliefs that exist elsewhere but fail to activate in risk contexts (e.g., “I want to stay healthy,” “I can cope without alcohol”). These unused beliefs provide anchors for adaptive inference chains.

In the second step, new inference chains must be conceptualized that (a) make harmful behavior less coherent and (b) strengthen predictions of alternative, goal-aligned behavior. For example: “Craving does not define who I am but my goals do” → “I have chosen not to drink” → “This decision is stronger than craving” → “I will abstain in this situation.” The effectiveness of this step depends on aligning interventions with self-concept and prior expectations, ensuring the new chain better minimizes expected surprise than the harmful one. It is crucial to recognize that prior behavior (and the potential label of having a specific mental disorder) as well as current thoughts and feelings may serve as a relevant argument for related self-concept beliefs. Therefore, behavior process chains must be considered that outweigh these arguments, providing stronger support for alternative behavior.

Finally, the intervention phase should integrate multiple strategies to support these inference chains. Inference nudging interventions can tailor risk situations to promote relevant inferences. For instance, reminders in drinking contexts (e.g., on the fridge) can highlight that drinking to reduce stress is less effective than alternative strategies, thereby promoting the inference that healthier actions better align with personal goals. Importantly, however, this technique should always be used in reference to the individual’s current beliefs, ensuring that the inference chain effectively reduces expected surprise (e.g., the reminder should not be used if individuals typically infer that relaxation is ineffective or unattainable to them).

One could also implement inference training techniques that strengthen the prediction that alternative actions will be enacted in future situations, such as restructuring the environment so that engaging in relaxation exercises consistently results in positive outcomes and discussing how these contingencies relate to personal goals and are likely to motivate behavior change. Practice-based approaches may be implemented to support this. Techniques such as ABC-training (which involves exposure to antecedent contexts and making choices between alternative actions and alcohol drinking in light of goal-related consequences; Wiers et al., 2020) can support this process. This type of training may impact relevant inferences and alcohol consumption (Van Dessel et al., 2023; Pan et al., 2024), but its success hinges on individuals inferring that they will also apply the practiced behaviors in everyday contexts. To support generalization, training can be combined with implementation intentions (e.g., 'if I feel stressed, then I will drink water'), encouragement to practice these if-then inferences in daily life (Adriaanse et al., 2011; Bélanger-Gravel et al., 2013), and discussions about the effectiveness of the training (Linthout et al., 2025).

Together, these strategies illustrate how GDPP provides a principled framework for reconfiguring inference chains, showing not only why harmful alcohol consumption occurs, but also how interventions can systematically redirect predictions toward adaptive behavior.

Applied Case: Depressive Behavior

From a GDPP perspective, depression can be understood as a pattern of withdrawal and passivity that is maintained by a network of hierarchical beliefs. A common form of harmful behavior in depression is the persistent avoidance of goal-directed action, such as applying for a job or engaging in social contact. This avoidance may seem irrational at first glance, but within the individual's generative belief system, it can offer a coherent and low-surprise resolution to conflicting predictions and entrenched self-beliefs (Miller et al., 2022).

Consider a person who, despite wanting a job, consistently avoids applying. When confronted with a job opportunity, they may activate the belief that they should have a job, which generates a prediction error. Yet instead of updating toward action, the person may infer withdrawal because it aligns with high-level beliefs and better reduces expected surprise within their current belief network.

The first step in a GDPP-based intervention is to identify the inference chains that support the harmful behavior. The practitioner might guide the person to reflect on the behavior, its antecedents and its consequences, including associated behaviors such as feelings of sadness and negative thoughts (e.g., “Things will never get better for me”). Feelings of distress may also be discussed to reveal competing beliefs that introduce disorder in the belief system. Self-concept beliefs are of particular interest, given their central role in organizing inference chains. Finally, the practitioner can also consider beliefs that often underlie avoidance as observed in prior research, such as the belief that one will not strive for desired outcomes because they are unobtainable (e.g., due to generalization from prior failure-related experiences: Boddez et al., 2022).

Following this identification of beliefs based on clinical observations, discussions, and prior literature, hypotheses about the functional role of these beliefs can be tested, for instance, by examining whether they help explain the persistence of the behavior across contexts. This process allows for the mapping of likely inference chains that minimize surprise within the current belief network, such as: “I want to apply for a job” → “I have failed before” → “I will feel worse if I try and fail again” → avoidance. By mapping out these inference steps, the practitioner makes visible both the coherence and the disorder within the belief system and identifies possible entry points for change.

The second step is to help the person construct alternative inference chains that support goal-directed behavior while preserving coherence within their belief system. The practitioner might highlight prior experiences that contradict the global belief of inevitable failure or explore valued goals that are still attainable. This may help the person articulate beliefs that are better aligned with their goals and result in the conceptualization of new target inference chains, such as: “I want to apply for a job” → “Trying to apply is aligned with my goals and beliefs” → “I am able to apply for a job” → applying for a job. These alternative inferences are likely to reduce surprise if they resonate with the person’s values and self-concept, offering a viable and coherent alternative to avoidance.

The final step involves reinforcing the selected inference chain by helping the person enact and apply it across relevant contexts. This step aims to shape a belief network that is coherent and strengthens the predictive value of adaptive inferences. Therapeutic dialogue may challenge the internal coherence of avoidance-based beliefs and support the development of alternative beliefs that align with the evolving self-concept. This can be complemented with mental simulation, real-life enactment of new inferences, and consistent feedback on their positive consequences. Interventions may also incorporate explicit scaffolds that help the person reconnect with the target inference in difficult moments. For example, small steps toward action (such as drafting a job application) might be combined with brief self-instruction (e.g., “This is hard, but trying aligns with who I want to be”) or visual reminders of meaningful goals. These reinforcements increase the coherence and predictability of the new belief chain across situations.

Whereas Cognitive Behavioral Therapy (CBT) often promotes belief change, GDPP clarifies how, why, and when this is likely to succeed: adaptive beliefs must become more predictive of behavior than maladaptive ones by reducing expected surprise. While CBT is considered one of the most effective treatments for depression (Fennell, 2012), relapse rates remain high (Lorimer et al., 2019). The GDPP framework suggests that outcomes may improve by more explicitly targeting underlying inferences and self-concept beliefs in relation to expected surprise. This can be achieved by practicing the application of adaptive beliefs in context. For instance, repeated practice to act upon the belief “Avoidance and associated negative thoughts and feelings are understandable but no longer fit with how I see myself, so they don't determine what I do" may gradually reshape expectations and reduce the grip of passivity (Vermeersch, 2023). This resembles cognitive bias modification (CBM), which has recently been integrated into CBT yet often lacks a coherent theoretical foundation (Vrijsen et al., 2018). GDPP offers such a framework and explains why and when such techniques may be effective: they may succeed when they reorganize belief networks in a way that reduces expected surprise and reinforces predictions aligned with important goals.

Conclusion

This paper has introduced the goal-directed predictive processing (GDPP) framework, an integrative approach grounded in recent theoretical advancements and robust empirical findings. The framework is designed to equip practitioners with theory-based guidance for understanding and modifying harmful behavior. By relying on streamlined assumptions with high heuristic value, it can offer a practical and adaptable tool for behavior change interventions while maintaining flexibility for future refinements.

This is the first detailed exposition of the GDPP framework. Our aim was not to provide a comprehensive review of or comparison with all theoretical models of harmful behavior, but to demonstrate how insights from predictive processing and goal-directed theories can be integrated into a cohesive framework relevant for practice. Rather than aligning with every aspect of the underlying theories, the GDPP framework selectively integrates components that offer substantial heuristic value and may be most relevant for practical interventions. By prioritizing mental processes over neural-level explanations, the framework remains accessible and actionable for practitioners, acknowledging the inherent challenges of directly linking cognitive processes to neural mechanisms in ways that inform intervention (Niv, 2021). As neuroscience and behavioral science advance, future versions of GDPP may incorporate neural insights to further expand its scope.

Because the GDPP framework aims to integrate the most robust empirical evidence, it is inherently a work in progress. Ongoing research is crucial to validate its predictive power and refine its assumptions based on emerging findings and collaborative efforts (see [link blinded for review]). To support this, GDPP-based interventions should be empirically evaluated. The framework enables the generation of testable predictions about which techniques will most effectively reduce harmful behavior. Evaluations across behavioral domains will be essential for assessing its generalizability and practical value. This continuing development will be essential for maximizing its utility in guiding effective, theory-based behavior change practices. In doing so, GDPP offers practitioners a principled framework for translating the most robust scientific insights into effective behavior change.

Acknowledgements

We thank Agnes Moors, Tilia Linthout, and Reinout Wiers for helpful feedback on an earlier draft of this manuscript.

Statements and declarations

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article'

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Ghent University Methusalem grant 01M0020 and Ghent University Special Research Fund grant BOF/STA/202202/004.

References

Abramowitz, J. S., Deacon, B. J., & Whiteside, S. P. (2019). Exposure therapy for anxiety: Principles and practice. Guilford Publications.

Adriaanse, M. A., Vinkers, C. D. W., De Ridder, D. T. D., Hox, J. J., & De Wit, J. B. F. (2011). Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. *Appetite, 56*, 183-193. doi:10.1016/j.appet.2010.10.012

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50(2),* 179–211. doi:10.1016/0749-5978(91)90020-T

Albarracín, D., Fayaz-Farkhad, B., & Granados Samayoa, J.A. (2024). Determinants of behaviour and their efficacy as targets of behavioural change interventions. *Nature Reviews Psychology, 3*, 377–392. doi:10.1038/s44159-024-00305-0

Bagozzi, R.P., & Pieters, R. (1998). Goal-directed Emotions. *Cognition and Emotion, 12*(1), 1-26. doi:10.1080/026999398379754

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215. doi:10.1016/0146-6402(78)90002-4

Baumeister, R. F., Vohs, K. D., Nathan DeWall, C., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and social psychology review, 11*(2), 167-203. doi:10.1177/1088868307301033

Beck, J. S. (2011). Cognitive behavior therapy: Basics and beyond. Guilford Press.

Bélanger-Gravel, A., Godin, G., & Amireault, S. (2013). A meta-analytic review of the effect of implementation intentions on physical activity. *Health Psychology Review, 7(1),* 23–54. doi:10.1080/17437199.2011.560095

Bem, D. J. (1972). Self-Perception Theory. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 6, pp.1-62). New York: Academic Press.

Berkman, E. T., & Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (social) psychological theory. *Perspectives on Psychological Science, 16*(4), 864–874. doi:10.1177/1745691620969650

Biglan, A. (2015). The nurture effect. How the science of human behavior can improve our lives and our world. Oakland, CA : Harbinger.

Boddez, Y., Van Dessel, P., & De Houwer, J. (2022). Learned helplessness and its relevance for psychological suffering: a new perspective illustrated with attachment problems, burn-out, and fatigue complaints. *Cognition & Emotion, 36*(6),1027-1036. doi:10.1080/02699931.2022.2118239

Buabang, E., Köster, M., Hogarth, L., & Moors, A. (2024). Poor reliability and validity of habit effects in substance use and novel insights from a goal-directed perspective. doi:10.31234/osf.io/79ykb

Carver C. S., & Scheier M. F. (1990). Origins and functions of positive and negative affect: a control-process view. *Psychological Review, 97*(1), 19.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of educational research, 63*(1), 1-49. doi:10.3102/003465430630010

Clark, A. (2013). Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behavioral and Brain Sciences, 36*, 181–204. doi:10.1017/S0140525X12000477

Clark, A. (2016). Surfing Uncertainty: Prediction, Action, and the Embodied Mind. New York: Oxford Academic.

Clark, A. (2020). Beyond desire? Agency, choice, and the predictive mind. Australasian Journal of Philosophy, 98(1), 1–15. doi:10.1080/00048402.2019.1602661

Cleeremans, A. (2011). The Radical Plasticity Thesis: How the brain learns to be conscious. *Frontiers in Psychology, 2*, 1-12. doi:10.3389/fpsyg.2011.00086

Cotman, C. W., & McGaugh, J. L. (2014). Behavioral neuroscience: An introduction. Academic Press.

David, D., Cristea, I., & Hofmann, S.G. (2018). Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy. *Frontiers in Psychiatry, 29*, 9:4. doi:10.3389/fpsyt.2018.00004

Deci, E. L., & Ryan, R. M. (2012). Self-Determination Theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), Handbook of Theories of Social Psychology (pp. 416-437, Vol. 1). Thousand Oaks, CA: Sage. doi:10.4135/9781446249215.n21

De Graaf, N. D., & Wiertz, D. (2019). Societal problems as public bads. Routledge.

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and personality psychology compass, 8*(7), 342-352. doi:10.1111/spc3.12111

De Houwer, J., & Hughes, S. (2020). Learning psychology: An introduction from a functional-cognitive perspective. Macmillan Education.

De Houwer, J., Finn, M., Boddez, Y., Hughes, S., & Cummins, J. (2024). Relating different perspectives on how outcomes of behavior influence behavior. *Journal of the*

Experimental Analysis of Behavior, 121(1), 123-133. doi:10.1002/jeab.887

Fennell, M. (2012). Cognitive behaviour therapy for depressive disorders. In: Gelder M, Andreasen N, Lopez-Ibor J, Geddes J, editors. New Oxford Textbook of Psychiatry. New York: Oxford University Press, pp. 1304–12.

Fisher, W. W., Piazza, C. C., & Roane, H. S. (2021). Handbook of applied behavior analysis (2nd ed.). Guilford Press

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11:127–38. doi: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.

Gardner, H. (1987). The mind's new science: A history of the cognitive revolution. New York: Basic Books.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). Heuristics and biases: The psychology of intuitive judgment. Cambridge University Press.

Hagger, M. S., Cameron, L., Hamilton, K., Hankonen, N., & Lintunen, T. (2020). The handbook of behavior change. New York, NY: Cambridge University Press.

Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie, 38*(1), 7–16. doi:10.1024/0044-3514.38.1.7

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science, 1*, 1-16. doi:10.1016/j.jcbs.2012.09.004

Hayes S.C., Luoma, J.B., Bond, F.W., Masuda, A., Lillis, J. (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behavior Research and Therapy, 44*, 1–25. doi:10.1016/j.brat.2005.06.006

Helmholtz, H. (1962). Handbuch der physiologischen optic. J. P. C. Southall (Ed.), Vol. 3. New York: Dover.

Hogarth, L. (2020). Addiction is driven by excessive goal-directed drug choice under negative affect: Translational critique of habit and compulsion theory. *Neuropsychopharmacology, 45,* 720–735. doi:10.1038/s41386-020-0600-8

Hohwy, J. (2013). The Predictive Mind. Oxford University Press.

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk.

Econometrica, 47(2), 263-291. doi:10.2307/1914185

Köster, M., Moors, A., De Houwer, J., Ross-Hellauer, T., Van Nieuwerburgh, I., & Verbruggen, F. (2021). Behavioral Reluctance in Adopting Open Access Publishing: Insights From a Goal-Directed Perspective. *Frontiers in Psychology, 12*, 649915. doi:10.3389/fpsyg.2021.649915

Kirsch, I. (1990). Changing Expectations: A New Method for Predicting Therapeutic Effect. University of Minnesota Press.

Kube, T., & Rozenkrantz L. (2021). When Beliefs Face Reality: An Integrative Review of Belief Updating in Mental Health and Illness. *Perspectives on Psychological Science, 16*(2), 247-274. doi:10.1177/1745691620931496

Kruglanski, A. W., & Szumowska, E. (2020). Habitual behavior is goal driven. *Perspectives on Psychological Science, 15(5),* 1256–1271. doi:10.1177/1745691620917676

Leshner A.I. (1997). Addiction is a Brain Disease, and it Matters. *Science New Series, 278*, 45–7. doi:10.1176/foc.1.2.190

Linthout, T., Van Dessel, P., & Boddez, Y. (2025). Targeting Predicted Actions to Shift Impulsive Behavior. *OSF preprints.*

Locke, E. A., & Latham, G. P. (1990). A Theory of Goal Setting & Task Performance. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Locke, E.A., & Latham, G.P. (2006). New Directions in Goal-Setting Theory. *Current Directions in Psychological Science, 15*, 265–268. doi:10.1111/j.1467-8721.2006.0044

Lorimer, B., Delgadillo, J., Kellett, S., & Brown, G. (2019). Exploring relapse through a network analysis of residual depression and anxiety symptoms after cognitive behavioural therapy: A proof-of-concept study. *Psychotherapy Research, 30*(5), 650-661. doi:10.1080/10503307.2019.1650980

Lovibond P. F. (2006). Fear and avoidance: an integrated expectancy model, in Fear and Learning: From Basic Processes to Clinical Implications, eds Craske M. G., Hermans D., Vansteenwegen D. Washington, DC: American Psychological Association, 117–132.

Marteau, T. M. (2018). Changing minds about changing behavior. *The Lancet, 391*(10116), 116-117. doi:10.1016/S0140-6736(17)33324-X

Mertens, S., Herberz, M., Hahnel, U. J. J. & Brosch, T. The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences, 119*, e2107346118. doi:10.1073/pnas.2107346118

Michie, S., & West, R. (2013). Behaviour change theory and evidence: A presentation to government. *Health Psychology Review, 7* , 1–22. doi:10.1080/17437199.2011.649445

Milkman, K. L. et al. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proceedings of the National Academy of Sciences, 119,* Article e2115126119. doi:10.1073/pnas.2115126119.

Miller, M., Kiverstein, J., & Rietveld, E. (2022). The Predictive Dynamics of Happiness and Well-Being. *Emotion Review, 14*(1), 15-30

Moors, A., Boddez, Y., & De Houwer, J. (2017). The power of goal-directed processes in the causation of emotional and other actions. *Emotion Review, 9*, 310-318. doi:10.1177/175407391666959

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology, 67*.

Moors, A. (2019). Towards a goal-directed account of weak-willed behavior. *The Brains Blog*. doi:10.23668/psycharchives.5690

Moors, A. (2022). Demystifying emotions: A typology of theories in psychology and philosophy. Cambridge University Press.

Moors, A., Van de Cruys, S., & Pourtois, G. (2021). Comparison of the determinants for positive and negative affect proposed by appraisal theories, goal-directed theories, and predictive processing theories. *Current Opinion in Behavioral Sciences, 39*, 147–15. doi:10.1016/j.cobeha.2021.03.015

Munakata, Y., Placido, D., & Zhuang, W. (2023). What's next? Advances and challenges in understanding how environmental predictability shapes the development of cognitive control. *Current Directions in Psychological Science, 32*(6), 431–438. doi:10.1177/09637214231199102

Niv, Y. (2021). The primacy of behavioral research for understanding the brain. *Behavioral Neuroscience, 135*(5), 601–609. doi:10.1037/bne0000471

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pan, T., Szpak, V., Laverman, J., Van Dessel, P., Bovens, R., Larsen, H., & Wiers, R.W. (2024). ABC-Training for Alcohol Use During a Voluntary Abstinence Challenge: A Randomized Controlled Trial. International Journal of Mental Health and Addiction. https://doi.org/10.1007/s11469-024-01409-7

Pierce, W. D., & Cheney, C. D. (2018). Behavior analysis and learning: A biobehavioral approach (6th ed.). New York, NY: Routledge.

Petry, N. M. (2011). Contingency management: what it is and why psychiatrists should want to use it. *Psychiatrist, 35*(5), 161-163. doi:10.1192/pb.bp.110.031831

Parr T., Pezzulo, G., & Friston K. (2022). Active Inference: The Free Energy Principle in Mind, Brain, and Behavior. MIT Press.

Pezzulo, G., Rigoli, F. Friston, K. (2018). Hierarchical Active Inference: a Theory of Motivated Control. *Trends in Cognitive Sciences, 22*(4), 294-306. doi:10.1016/j.tics.2018.01.009

Roefs, A., Fried, E.I., Kindt, M., Martijn, C., Elzinga, B., Evers, A. W. M, Wiers, R. W., Borsboom, D., & Jansen, A. (2022) A new science of mental disorders: Using personalised, transdiagnostic, dynamical systems to understand, model, diagnose and treat psychopathology. *Behaviour Research Therapy, 153*, 104096. doi:10.1016/j.brat.2022.104096

Rosenberg, M. (1979). Conceiving the self. Appleton-Century-Crofts.

Skinner, B. F. (1953). Science and human behavior. Macmillan.

Sprevak, M., & Smith, R. (2023). An introduction to predictive processing models of perception and decision-making. *Topics in Cognitive Science*. Advance online publication. https://doi.org/10.1111/tops.12704

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. Yale University Press.

Thorndike, E. L. (1905). The elements of psychology. New York, NY: A. G. Seiler.

Vanaken, L., Boddez, Y., Bijttebier, P., & Hermans, D. (2021). Reasons to remember: A functionalist view on the relation between memory and psychopathology. *Current Opinion in Psychology, 41*, 88-95.

Van de Cruys, S. (2017). Affective value in the predictive mind. In: Metzinger, T. K., & Wiese, W. (Eds.), Philosophy and Predictive Processing, MIND Group, 1-21.

Van de Cruys, S., & Van Dessel, P. (2021). Mental distress through the prism of predictive processing theory. *Current Opinion in Psychology, 41*, 107–112. doi:10.1016/j.copsyc.2021.07.006.

Van Dessel, P., Boddez, Y., & Hughes, S. (2022). Nudging societally relevant behavior by promoting cognitive inferences. *Scientific Reports, 12*, 9201. doi:10.1038/s41598-022-12964-1

Van Dessel, P., Cummins, J., & Wiers, R.W. (2024). ABC-Training As a New Intervention For Hazardous Alcohol Drinking: Two Proof-of-Principle Randomized Pilot Studies. *Addiction.* doi:10.1111/add.16271

Van Dessel, P., Fedeli, F., Zogmaister, C., & Boddez, Y. (2025). From Goal-Directed Inference Chains to Action: An Integrative Framework for Promoting Sustainable Behaviour. *OSF* *preprints.* https://doi.org/10.31219/osf.io/qr64b_v1

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). Consequence-Based Approach-Avoidance Training: A New and Improved Method for Changing Behavior. *Psychological Science, 29*, 1899-1910. doi:10.1177/0956797618796478

Vanderschoot, T., & Van Dessel, P. (2022). EMDR Therapy and PTSD: A Goal-Directed Predictive Processing Perspective. *Journal of EMDR Practice and Research*. doi:10.1891/EMDR-2022-0009

Vermeersch, C. (2023). Training van contrasterende overtuigingen bij depressieve klachten: Kunnen we depressieve klachten reduceren door herhaaldelijke toepassing van contrasterende overtuigingen? [Master's thesis, Universiteit Gent]. https://lib.ugent.be/catalog/rug01:003146179

Vermeir, I., Weijters, B., De Houwer, J., Geuens, M., Slabbinck, H., Spruyt, A., Van Kerckhove, A., Van Lippevelde, W., De Steur, H., & Verbeke, W. (2020). Environmentally Sustainable Food Consumption: A Review and Research Agenda From a Goal-Directed Perspective. *Frontiers in Psychology, 11*, 1603.

Vrijsen, J.N., Fischer, V., Müller, B.W., Scherbaum, N., Becker, E.S., Rinck, M., & Tendolkar, I. (2018). Cognitive bias modification as an add-on treatment in clinical depression: Results from a placebo-controlled, single-blinded randomized control trial. *Journal of Affective Disorders, 238*,342-350.

Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press.

Wiers, R., Van Dessel, P., & Kopetz, C. (2020). ABC-training: a new theory-based form of cognitive bias modification to foster automatization of alternative choices in the treatment of addiction and related disorders. *Current Directions in Psychological Science, 29,* 499-505. doi:10.1177/096372142094950

Wiers, R.W., Rinck, M., Kordts, R., Houben, K., & Strack, F. (2010). Re-training automatic action-tendencies to approach alcohol in hazardous drinkers. *Addiction, 105*, 279–287. doi:10.1111/j.1360-0443.2009.02775.x

Wiers, R. W., Gladwin, T. E., Hofmann, W., Salemink, E., Ridderinkhof, K. R. (2013). Cognitive bias modification and cognitive control training in addiction and related psychopathology: mechanisms, clinical perspectives, and ways forward. *Clinical Psychological Science, 1*(2),192–212. doi:10.1177/2167702612466547