Running head: GOAL-DIRECTED EMOTIONAL AND OTHER ACTION

# The power of goal-directed processes in the causation of emotional and other actions

Agnes Moors[1,2], Yannick Boddez[1], & Jan De Houwer[2]

[1] KU Leuven

[2] Ghent University

Corresponding author: Agnes Moors, KU Leuven – University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium, Tel. +32/16373279, agnes.moors@kuleuven.be

Author Note: Preparation of this paper was supported by Research Programme G.0223.13N of the Research Foundation - Flanders (FWO), Research Fund of KU Leuven (GOA/15/003 and PF/10/005), Interuniversity Attraction Poles grant of the Belgian Science Policy Office (P7/33), and Methusalem Grant BOF09/01M00209 of Ghent University.

Abstract

Standard dual process models in the action domain postulate that stimulus-driven processes are responsible for suboptimal behavior because they take them to be rigid and automatic and therefore the default. We propose an alternative dual process model in which goal-directed processes are the default instead. We then transfer the dual process logic from the action domain to the emotion domain. This reveals that emotional action tendencies are often attributed to stimulus-driven processes. Our alternative model submits that emotional action tendencies can also be caused by goal-directed processes. We evaluate the type of empirical evidence required for validating our model and we consider implications of our model for behavior change, encouraging strategies focused on the expectancies and values of action outcomes.

*Keywords:* action, emotion, dual process, goal-directed, stimulus-driven, automatic

## The power of Goal-directed Processes in the Causation of Emotional and Other Actions

People often engage in behavior that is not in their best interest, so-called irrational or suboptimal behavior. For instance, they engage in unhealthy behavior such as smoking; they commit action slips such as taking a wrong habitual route; or they become aggressive even when they risk a high cost (e.g., retaliation or ruining a relationship). Smoking and taking the wrong route have been framed as the result of (bad) habits, and are typically cited in the domain of action research. Costly aggression falls in the domain of emotion research, and people committing to this kind of aggression have been said to be in the grip of an emotion (anger). To explain dissociations between optimal and suboptimal behavior, theorists in both the domains of action and emotion have come up with dual process models in which goal-directed processes are responsible for much of our optimal behavior and stimulus-driven processes are more likely to lead to suboptimal behavior.

The aim of this article is two-fold. The first aim is to challenge standard dual process models in the action literature in which stimulus-driven processes are assumed to be the driving force of suboptimal behavior. The main source of this assumption, in our view, is the postulate that stimulus-driven processes are rigid but automatic and therefore the default type of process whereas goal-directed processes are flexible but only operate under special conditions. Based on a rejection of this postulate, we present an alternative dual process model, which takes goal-directed rather than stimulus-driven processes to be the default type of process. This amounts to reducing the explanatory territory of stimulus-driven processes and expanding that of goal-directed ones.

The second aim is to extend our analysis from the action domain to the emotion domain. This exercise reveals that the causation of emotional action tendencies is often conceptualized in terms of stimulus-driven rather than goal-directed processes. Here too then, stimulus-driven processes are considered the default type of process whereas goal-directed

processes are confined to a secondary, regulatory role. The alternative model that we put forward, however, proposes to also consider goal-directed processes as a potential cause of initial emotional action tendencies.

After our theoretical analysis, we evaluate some of the existing empirical work in light of standard dual process models and our alternative model, and we make suggestions for future empirical research aimed at pitting both models against each other more directly. We close by elaborating on the implications of our alternative model for behavior change in clinical practice and society.

## Dual Process Models in the Action Domain

Dual process models in the action domain distinguish between goal-directed and stimulus-driven processes. A goal-directed process assesses the utility of one or more action options. The utility of an action option is based on (a) a representation of the *value* of the outcome of the action, which in turn, depends on the motivational state or desire of the organism, and (b) a representation of the contingency between the action and the outcome, also called the *expectancy* or the belief that the action will lead to the outcome (Heyes & Dickinson, 1993). The action option with the highest utility activates its corresponding action tendency (i.e., goal to act), and this action tendency may or may not translate in an overt action. If it results in an overt action, that action is said to be under the control of its expected and desired outcomes (i.e., expected goals) and is called an instrumental action. The sequence from stimulus input (S) over representations to response output (R) can be formalized as follows (with the content of representations inside square brackets, and observable inputs and outputs outside these brackets)[1]:

[1] The term stimulus may cover any aspect of the environment that is taken as the input of a process. The term action is understood broadly as overt behavior and motor responses. The term outcome refers to a stimulus that is caused by an action. The term goal refers to the representation of a valued or desired outcome.

In this manuscript, we adopt the common practice to define instrumental behavior and reactions at a mental level of analysis, that is, in terms of the mental processes causing them. Note that these concepts can also be defined at an observable or functional level of analysis (see De Houwer, 2011): Instrumental behavior can be

$$S \rightarrow [S{:}R1\text{-}O1^{v}]\ [S{:}R2\text{-}O2^{v}] \rightarrow [R1] \rightarrow R1$$

The first two representations contain the utilities of two action options (R1 and R2), each connected to a unique valued outcome ($O1^{v}$ and $O2^{v}$) to a certain degree. The third representation ([R1]) contains an action, and is called an action tendency or goal to act. More complex formulas are possible with more action options connected to multiple action outcomes. But the minimal requirement for a process to count as goal-directed is that at least one action option is represented together with its expectancy for one valued outcome.[2]

It is insightful to embed the goal-directed process in a cycle in line with cybernetic models of action control (Carver & Scheier, 1990; Ridderinkhof, this issue). Each cycle starts with the comparison between a stimulus (actual state) and a *first* goal (i.e., representation of a valued outcome). If a discrepancy is detected, the organism strives to reduce the discrepancy (i.e., a *second* goal), either by acting (i.e., assimilation), by choosing a different first goal (i.e., accommodation), or by biasing the interpretation of the outcome (i.e., immunization; Brandtstädter & Rothermund, 2002). The utility of acting in general and of specific action options determines whether and which action tendency (i.e., *third* goal) will be activated. If an action tendency is activated and translated into overt action (i.e., assimilation), the outcome of this action feeds back to the comparator where it constitutes the stimulus input to the next cycle. New cycles are run until there is no discrepancy left. Cybernetic models further assume that people must balance multiple goals, which implies that several cycles may run in parallel. Moreover, they assume a hierarchical organisation such that cycles for lower-order goals are embedded in cycles for higher-order goals. That is, the discrepancy with a lower-order goal can be considered as a stimulus that is itself discrepant with a higher-order goal. The goal to keep multiple lower-order goals in balance is an example of such a

defined as behavior that is a function of its consequences, and a reaction (or habit) as behavior that is a function of the antecedent stimulus. Both types of definitions may yield non-overlapping sets of behavior.

[2] Note that on this narrow definition of goal-directed, the activation of a goal to act via a goal priming procedure does not count as a goal-directed process.

higher-order goal (see also Carver & Scheier, 1990; Gross, 2015; Saunder, Milyavskaya, & Inzlicht, 2015; Webb et al., 2012).

In contrast to the goal-directed process, a stimulus-driven process involves the activation of a pre-existing association between the representation of a specific stimulus and the representation of a specific action. The stimulus representation can be made up of a single feature or a collection of features, and it can be activated by any stimulus that shares features with it. The action representation can be considered as an action tendency[3], which may or may not translate in overt action. Here, the action is said to be under the control of the antecedent stimulus, and is sometimes called a re-action instead of an action. The sequence of events can be formalized as follows:

$$S \rightarrow [S\text{-}R] \rightarrow R.$$

The representations in both goal-directed and stimulus-driven processes can come about via multiple routes. The representation of an $S{:}R\text{-}O^{v}$ link can be installed via an operant conditioning procedure in which a response to a stimulus is followed by an outcome (e.g., a person may touch the hot stove and feel pain), but it may also be installed via verbal instructions, observation, or logical inference (Heyes & Dickinson, 1993).

The representation of an S-R link can also come into existence via multiple routes. Some S-R links are innate: An unconditional stimulus (US, e.g., a loud noise) is wired together with an unconditional response (UR, e.g., to startle) from birth. Other S-R links are acquired via an overtrained operant conditioning procedure: When a response to a stimulus is extensively followed by the same outcome, the representation of the outcome is assumed to be no longer activated or even erased. Such a response is called a habit. Another way to install a S-R link is via an associative learning procedure: The mere co-occurence of a

[3] Here we follow James (1890) in that we do not make a distinction between a *goal* representation of an action and a *mere* representation of an action. James (1890) suggested that all representations of actions are inherently impulsive: Once activated they translate into overt action unless they are overpowered by a competing representation or physical circumstances.

stimulus and a response is supposed to lead to the binding of their representations (Hebb, 1949).[4] Still another route that has been proposed for the installation of an S-R link is the formation of an implementation intention (Gollwitzer & Sheeran, 2006) also called the prepared reflex (Hommel, 2000). An implementation intention is a goal to engage in a concrete action (e.g., leave the party) upon the occurrence of a concrete stimulus (e.g., when the clock sounds at midnight). It thus takes the form of an S-R link. Once the S-R link is put into place, the mere presence of the stimulus should suffice to activate this link and produce a response. Several findings, however, suggest that the intention has to remain active for the S-R link to produce a response (e.g., Sheeran, Webb, & Gollwitzer, 2005). On this interpretation, the stimulus-driven process is not merely preceded by, but remains embedded in a goal-directed process.

Standard dual process models assume that stimulus-driven processes are the driving force of suboptimal behavior. This assumption is rooted in the postulate that there is a trade-off between optimality and automaticity (Strack & Deutsch, 2004). Optimality refers to the degree of goal satisfaction that a process can bring about; automaticity refers to the number of conditions a process requires for its operation (e.g., time, attention, the intention to engage in the process; Moors & De Houwer, 2006). This trade-off is linked with the degree of complexity and hence flexibility/rigidity of the processes. Goal-directed processes are more complex, which makes them less automatic, but in return more flexible and therefore more likely to produce optimal behavior. Stimulus-driven processes are more simple, which makes them more automatic, but in return more rigid and therefore more liable to produce suboptimal actions in some cases (depending on the suitability of the pre-existing S-R link for the current occasion).

[4] An associative learning procedure (i.e., pairings of stimuli) is typically part of an operant learning procedure, but it can also occur in isolation, that is, without the presentation of an outcome.

The implication is that when operating conditions are poor (e.g., there is little time or attention), the organism is entirely at the mercy of stimulus-driven processes, which often result in suboptimal behavior. But even when operating conditions are ample, it is assumed that goal-directed processes can at best jump in at a later stage to regulate (refine or correct) the action tendency that was initially produced by a stimulus-driven process (e.g., Wood & Neal, 2007). Some standard dual process models propose a default-interventionist architecture in which the stimulus-driven process is the default process, whereas goal-directed processes only occur under special conditions (Evans & Stanovich, 2013; Wood & Rünger, 2016). Other standard dual process models propose a parallel-competitive architecture in which both types of processes constantly operate in parallel, but compete to determine behavior (Daw, Niv, & Dayan, 2005; Lee, Shimojo, & O'Doherty, 2014). Even in the latter models, however, the stimulus-driven process is expected to often win the competition because it is faster to produce a response.

## Alternative Dual Process Model

Like standard dual process models in the action domain, we divide the realm of processes into goal-directed and stimulus-driven processes based on the content of the representations involved ([S:R-$O^v$] vs. [S-R]). Most standard dual process models, however, go further in that they map this content-based dichotomy onto other dichotomies (Keren & Schul, 2009; Moors, 2014a). In the dual process models discussed, the dichotomy of stimulus-driven vs. goal-directed is typically mapped onto the dichotomy of automatic vs. nonautomatic. We challenge this mapping and submit that goal-directed processes can be fairly automatic too. Although we do recognize that stimulus-driven processes are simple and therefore likely to survive poor operating conditions, we do not discard the possibility that goal-directed processes can also arise under these conditions. Three arguments add credence to this position. First, goal-directed processes range from simple to complex and the simpler

ones (involving fewer action options and outcomes) are less likely to suffer from a paucity of operating conditions than the more complex ones (involving multiple action options and outcomes). Precisely how much complexity is possible under poor operating conditions is a question that must be studied in empirical research. There are already indications that complex information integration can be fast, unintentional, and even unconscious (Bechara, Damasio, Tranel, & Damasio, 1997; Hassin, 2013; Mudrik, Faivre, & Koch, 2014). For instance, Bechara et al. (1997) showed that participants were able to choose options with higher utilities without being able to consciously report on the more advantageous choice. This indicates that people can make unconscious estimates about optimal decisions. Second, we argue that the goals in goal-directed processes can partly compensate for the lack of other conditions (Moors, 2016). Support for this argument comes from studies in which subliminal priming effects occurred for goal-relevant but not goal-irrelevant primes (Tapia, Breitmeyer, & Schooner, 2010; see reviews by Eitam & Higgins, 2010, and Moors, 2016). This suggests that the goal-relevant stimuli were processed despite of the fact that they were masked and little time was available for processing them. Third, we not only reject the mapping of the stimulus-driven vs. goal-directed dichotomy (based on the content of representations) onto the automatic vs. nonautomatic dichotomy (based on the number of necesary operating conditions), but also onto the associative vs. rule-based dichotomy (based on the types of operations acting on representations; Sloman, 1996). One of the reasons why authors may resist abandoning the mapping of the first two dichotomies is that they link both to this third dichotomy (e.g., Wood & Neal, 2007, p. 844). However, if one accepts that goal-directed processes can also rely on computationally less taxing associative processes, the possibility of them being automatic becomes more intuitive.

If goal-directed processes can be fairly automatic too, there must be a substantial number of cases in which goal-directed and stimulus-driven processes are both able to

operate in parallel. This is consistent with the parallel-competitive architecture of those standard dual process models that assume that both processes operate in parallel and then compete to determine behavior (Daw et al., 2005; Lee et al., 2014). Yet while these standard models assume that stimulus-driven processes should often win the competition (see above), we submit that goal-directed processes should win the competition instead. This is because goal-directed processes combine automaticity with optimality, whereas stimulus-driven ones only deliver automaticity. Thus, we submit that goal-directed processes should be the default process: The system should prioritize processes that are most likely to lead to optimal behavior unless it is hindered to do so. If one further assumes that the complete absence of goals is unlikely, there does not seem much room to be filled by stimulus-driven processes. We think their influence may be limited to (at least) two cases. A first case is when the goal to act was only specified at an abstract, superordinate level, leaving the specific stimulus-response mappings open. Here, the stimulus-driven process does not enter in competition with the goal-directed process so that it does not risk being overruled, and therefore gets to determine behavior. Take the case of an action slip in which a person takes a wrong habitual route that is blocked due to road works since a few days. It is possible that the superordinate goal to go to work was activated, but not the subordinate goals about the specific turns to take, thus leaving room for the habitual process to steer behavior[5]. This idea received support from studies on implementation intentions (Eder, 2011; Gollwitzer & Sheeran, 2006) in which it was shown that (presumably) stimulus-driven behavior (e.g., approaching positive stimuli) is better counteracted when the superordinate goal to counteract this behavior is supplemented with a subordinate goal (i.e., an implementation intention) that specifies a different S-R mapping (e.g., "if a red frame appears, avoid positive stimuli"). This indicates

[5] Note that in other examples of action slips, the behavior may or may not be stimulus-driven. Take the case of a person who jumps in her car to go shopping but ends up driving to work instead. This behavior may be attributed to an S-R link (being in the car activates driving to work), but it may also be attributed to an S:R-O link (being in the car activates the goal to be at work and the responses needed to get there).

that stimulus-driven processes can impact behavior if they are not counteracted by subordinate goals that are framed at the same level of abstraction as the stimulus-driven process. A second case in which the stimulus-driven process may determine behavior is when operating conditions (e.g., time, attention) are extremely poor so that even simple goal-directed processes may not run or may not be completed. Note that it is difficult to empirically determine whether an action triggered under poor conditions was caused by a stimulus-driven process or by an incomplete or malfunctioning goal-directed process. If time and attentional resources are scarce, a person may launch a goal-directed process but be forced to act before the process is completed.

In conclusion, the alternative view that we propose sees a smaller explanatory territory for stimulus-driven processes to the benefit of goal-directed ones. This implies that most of the competition does not occur between stimulus-driven and goal-directed processes, but instead between different goal-directed processes. Which process wins the competition is not only determined by complexity proportional to operating conditions, but also by the values and expectancies of outcomes.

Standard dual process models of action have documented various ways in which goal-directed and stimulus-driven processes can interface (e.g., Kotabe & Hoffmann, 2015; Wood & Neal, 2007; Wood & Rünger, 2016). For instance, stimulus-driven processes can result from overtrained goal-directed ones (cf. operant conditioning); they can be preceded or embedded in goal-directed ones (cf. implementation intentions); and they can interact when operating at the same time. By pointing at interfaces between goal-directed and stimulus-driven processes, authors acknowledge that behavior in daily life is often determined by a mixture of processes and that isolating either of them is difficult. In this way, they already go some way in reducing the territory of purely stimulus-driven processes. But talking about interfaces does not question the existence of the two processes in the interface, nor the

standard way of characterizing them (i.e., stimulus-driven as automatic, goal-directed as nonautomatic). The alternative model that we propose does not question the existence of both processes, but it does question the standard way of characterizing them. This leads to an even stronger reduction of the explanatory territory of stimulus-driven processes compared to standard models.

## Application to the Emotion Domain

In the present section, we analyze emotion theories through the goggles of the dual process models from the action domain. In the emotion domain, a contrast is often drawn between instrumental actions and emotional actions (Baumeister, 1997; Berkowitz, 1993). For instance, a father may raise his voice to make a child obey (i.e., an instrumental action) or he may perform the same behavior out of anger (i.e., an emotional action). Instrumental actions are thought to be caused by the same goal-directed processes as those invoked in the action literature. There is less consensus, however, about the processes causing emotional actions. In common sense explanations (and theories that try to vindicate them, e.g., Scarantino, this issue), emotional actions are said to be caused by emotions (e.g., aggressive behavior is caused by anger)–end of story. According to contemporary emotion theories in psychology, however, emotional actions are not caused by, but rather part of, emotions. These theories conceptualize emotions as episodes in which the processing of certain stimulus features gives rise to an action tendency. The action tendency activates physiological responses, and together they prepare and support overt behavior (i.e., whole body actions and facial/vocal expressions). Aspects of this episode seep into consciousness where they form the content of feelings. In these theories, then, an emotional action is one of several components of an emotional episode and is assumed to be caused by a process in which the *representation of certain stimulus features* activates an action tendency (i.e., the *representation of an action*) characteristic of an emotion. This fits our definition of a

stimulus-driven process. In some theories (e.g., affect program theories, Matsumoto & Ekman, 2009; network theories, Lang, 1994), the stimulus features giving rise to the action tendency are perceptual in nature (e.g., the form of a snake spurs the tendency to avoid). In other theories, however, the stimulus features are more abstract. This is the case for appraisal theories where the abstract features are called appraisals. Examples of appraisals are goal in/congruence (i.e., whether a stimulus is in/congruent with a person's goals) and high/low control (i.e., whether a stimulus is easy/difficult to control). Goal incongruence overlaps with the notion of a discrepancy between stimuli and goals in cybernetic models (cf. above), and high/low control indicates whether there are more/less action options with a high expectancy for reaching a valued outcome. Although this suggests partial overlap between the appraisal process and the goal-directed process, none of the current appraisal theories include an appraisal of the utility of *specific* action options afforded by a stimulus (i.e., affordances), and they do not hypothesize that the action tendency with the highest utility is always activated. Instead, specific patterns of appraisals (including goal in/congruence and high/low control) are hypothesized to give rise to specific action tendencies (e.g., to seek safety, attack, reject) that are seen as characteristic of specific emotions (e.g., fear, anger, disgust)[6]. Thus, the links between appraisal patterns and action tendencies put forward by appraisal theories do not count as goal-directed processes in the strict way defined here, but are better conceived of as sophisticated types of stimulus-driven processes. As such, they are less flexible than goal-directed processes, and therefore may lead to less optimal action tendencies in some situations. Take, for instance, the hypothesis of some appraisal theories that a stimulus appraised as goal incongruent and difficult to control leads to the tendency to flee or avoid, a tendency characteristic of fear (Roseman, 2013). In typical situations in which control is low

[6] Some appraisal theorists only link specific appraisals to specific action tendencies, without linking the latter to specific emotions (see Moors, 2014b).

(e.g., a conflict in which the person is weaker than the opponent), fleeing may be the optimal action, but there may be situations in which this is not true (e.g., if control is so low that fleeing is impossible). Likewise, fighting may be the optimal action in a typical high-control situation, but again, this may not always be the case (e.g., if the person has high control in the sense that she can flee but not fight). In sum, the processes proposed in several emotion theories can be classified as stimulus-driven ones, with stimulus features ranging from perceptual to abstract.

Crucially, as is the case in the action domain, dual process models in the emotion domain hold that complexity is positively related to flexibility and hence optimality, but inversely related to automaticity. Thus, goal-directed processes are seen as more flexible but in return less automatic, whereas the stimulus-driven processes held to be involved in emotional actions tendencies are seen as more rigid but in return more automatic. Indeed, it is assumed that raising one's voice to make someone obey is controlled, whereas raising it out of anger exemplifies a lack of control: The aggression spills out.

To ensure more flexibility in the relation between stimulus features and actions, the emotion theories discussed here allow for a scenario in which after the stimulus-driven process has generated a more or less abstract action tendency, a goal-directed process takes over that does take into account the concrete affordances of the situation. This process either refines an abstract action tendency into a more concrete one (e.g., fleeing can be done by running, swimming, or hiding under the carpet) or it corrects it (e.g., defensive fight instead of flight if fleeing is impossible). Such a goal-directed process is typically called a regulation process (see e.g., Tamir, 2009, for empirical evidence of the role of goal-directed processes in emotion regulation). This resembles the default-interventionist architecture that is typical of some standard dual process models in the action domain. We may thus conclude that a similar dual process logic underlies thinking in both the domains of action and emotion.

Application of our alternative dual process model to the emotion domain amounts to the proposal that goal-directed processes are not restricted to emotion regulation but should also play a role in the initial causation of emotional action tendencies. So far, only a handful of authors have explicitly considered this possibility (e.g., Bushman & Anderson, 2001; Eder & Hommel, 2013; Eder & Rothermund, 2013; Parkinson, 2008). The reason for this relative blind spot may be the widespread intuition that emotions have an irrational flavor. Theorists who define emotional actions by contrasting them with instrumental ones (cf. the dual process model outlined above) deny the possibility that emotional action tendencies can be caused by goal-directed processes. Yet if emotional actions are defined in an alternative way, the question becomes an empirical one. For instance, if emotional actions are defined as ensuing from action tendencies with *control precedence* (i.e., that take priority over other action tendencies) because these action tendencies, in turn, ensue from stimuli that are relevant for highly valued goals (Frijda, 2007), then there is no a priori reason to deny the role of goal-directed processes in the causation of emotional actions. On the contrary, if emotional episodes are concerned with highly goal-relevant events, it is only to hope that goal-directed processes will get into gear to prepare for the most optimal action. Theorists seem reluctant to abandon the irrational feature of emotions because they fear that it would turn emotions into cold decisions. Yet taking out the irrationality of emotions should not take out their heat: If they are defined as pertaining to highly goal-relevant events, their heat is preserved. Note that such an alternative definition of emotions relies on a gradual criterion: Episodes can be called more or less emotional. This squares well with emotion theories that deny a special status to emotions and their components (Russell, 2013). Instead of treating emotions as mysterious entities that influence cold cognitive processes involved in attention, perception, memory, and decision making, they can themselves be regarded as decision

processes, ones that rely on the very same mechanisms as non-emotional decision processes, but in which the stakes are higher.

## Empirical Validation of the Alternative Model

Research on operant learning, decision making, and motivation provides abundant support for the role of goal-directed processes in action causation. The challenge is to show that so-called instances of behavior caused by stimulus-driven processes are better explained by goal-directed processes. The evidence reviewed by habit researchers (e.g., Wood & Neal, 2007) and emotion researchers (Berkowitz, 1993; Matsumoto & Ekman, 2009) draws the image that stimulus-driven processes explain a substantial part of behavior. This seems to contradict the core assumption of our model that goal-directed processes are the default. In this section, we argue that there are problems with much of the evidence for stimulus-driven processes. The criteria used to diagnose stimulus-driven processes in this research range from weak to strong, but even strong criteria do not provide airtight guarantees about the underlying processes. This provides openings for future research to re-examine the role of goal-directed processes, and perhaps to redraw the image.

### Weak criteria

Weak criteria for the diagnosis of stimulus-driven processes are the frequency of behaviors and their (experienced) automaticity. In prediction research and intervention studies, habit strength is measured via self-report of the frequency of behaviors and/or of the degree to which they appear to be automatic (Verplanken & Orbell, 2003). But also in experimental studies, the stimulus-driven nature of habits and affective processes is often inferred from their automatic nature. Take the stimulus-driven hypothesis that positive vs. negative stimuli are sufficient to produce approach vs. avoidance tendencies (Chen & Bargh, 1999). This hypothesis received support from experiments in which positive vs. negative stimuli (i.e., pictures) led to faster instructed approach vs. avoidance responses (i.e., moving a

joystick or an avatar toward vs. away from the stimuli; see review by Krieglmeyer, De Houwer, & Deutsch, 2013). However, neither frequency nor automaticity are good indicators of the content of representations driving behavior. Brushing one's teeth may remain a choice, even if it is made everyday and carried out in a mindless way.

**Stronger criteria**

Stronger tests of the stimulus-driven vs. goal-directed nature of the processes that cause behavior are ones that do target the content of the representations involved such as the outcome revaluation test (i.e., the desire criterion test) and the contingency degradation test (i.e., the belief criterion test). If a change in the *value* of the expected outcome leads to a change in the response that previously led to the outcome, then it is concluded that a representation of the value of the outcome mediated the response (Adams & Dickinson, 1981). Likewise, if a change in the *contingency* between a response and an outcome leads to a change in the response that previously led to the outcome, then it is concluded that a representation of the contingency between the response and the outcome mediated the response (Balleine & Dickinson, 1998). Previous revaluation studies have shown that, after a *moderate* number of training trials in which a response is consistently followed by a specific positive outcome in the presence of a stimulus (S:R-posO), subsequent devaluation of the outcome (i.e., making it aversive by pairing it with a negative further outcome, negFO, or by changing the organism's motivational state, e.g., from hunger to satiation) leads to a decrement in the response that previously led to this outcome (Balleine & Dickinson, 1981). This finding is taken as evidence that the valued outcome was part of the representation causing the response ([S:R-posO]). After an *extensive* number of training trials, however, subsequent devaluation of the outcome no longer results in a response, a finding that is taken as evidence that the valued outcome was no longer part of the representation causing the response ([S-R]). Notably, overtraining does not result in the typical resistance-to-devaluation

effect when subjects (animals or humans) have the choice between two action options (Colwill & Rescorla, 1985). This indicates that behavior becomes habitual only under special conditions, such as extensive repetition and a lack of choice.

But even if revaluation and contingency degradation tests do show that a behavior does not change after changing the value and/or expectancy of its outcome, the conclusion that the underlying process was stimulus-driven is not airtight because it relies on null-findings (Heyes & Dickinson, 1993). An action that is insensitive to changes in the value or expectancy of one outcome (the manipulated one) may still be governed by the value and expectancy of another outcome (that was not manipulated). The goals driving behavior sometimes remain invisible (to the agent as well as to the researcher) because certain complexities are being overlooked. We discuss four complexities.

**Four complexities**

First, a single action may have multiple outcomes (benefits and costs), some of which may be hidden. Some cases attest to the dominance of the value of an outcome. For instance, aggression may come with the cost of retaliation, but it may be conducive to repairing one's damaged self-esteem and this may have a higher *value.* Likewise, people may endure extreme suffering to obtain a long-term outcome with a higher value for them (e.g., a martyr; Kopetz & Orehek, 2015). Other cases attest to the dominance of the expectancy of an outcome. For instance, a person taking drugs may be motivated by the immediate outcome of feeling high, and not deterred by the long-term outcome of poor health and job loss, possibly because the former is more vivid (Leigh, 1989) and therefore has a higher expectancy (cf. the availability heuristic; Tversky & Kahneman, 1974).

A second complexity is that multiple actions may lead to the same outcome. An individual may go for a suboptimal action option (one with a high cost) because there are no other options. For instance, a person may choose to repair self-esteem via costly aggression

because less costly action options (e.g., taking the opportunity to shine in public) are not available. In emotion research, aggression is sometimes measured by the amount of hot sauce or noise blasts delivered to an opponent but other actions options are often not presented (Gerber & Wheeler, 2009). Thus, it is difficult to determine whether participants chose it because of a tendency to be aggressive or a tendency to do anything.

A third complexity is that a single action may figure in various action-outcome links across contexts or times. If individuals keep records of previous action-outcome links that are no longer valid, these links may still sometimes be activated and drive behavior. For instance, smoking may breed popularity in one's teenage years, but cease to do so in one's adult years. It is possible that the old smoking-popular link was not overwritten but kept intact next to a new smoking-wrinkles link, and that both links are attached to different contexts (Gershman et al., 2013). In revaluation studies, the response decrement after moderate training may indicate that an old representation is overwritten by a new one, while the resistance-to-devaluation effect after extensive training may indicate that the old representation with the non-devalued outcome was activated because the test phase was more similar to the training phase than to the devaluation phase (e.g., because they share the same set of action options; for a related point, see Wood & Neal, 2007, Footnote 2).

A fourth complexity is that values and expectancies are not only determined by contingencies in the (recent) past, but also by meta-expectations. For instance, in contexts in which people believe there is a low contingency between actions and outcomes (e.g., in a gambling game), they may expect future outcomes to deviate from (instead of to replicate) previous outcomes (cf. the gamblers fallacy, e.g., Tversky & Kahneman, 1971). If this complexity is overlooked, actions erroneously may be taken to be insensitive to outcome expectancies.

Given that presumed evidence for stimulus-driven processes remains open to alternative interpretations, a fruitful approach would be to create variants of existing experiments in which the manipulation of values, expectancies, and/or meta-expectancies of presumed hidden goals impact on action tendencies and actions. Several labs have already taken this strategy. For example, Eder and Hommel (2013) review evidence obtained with the approach-avoidance task showing that the relation between positive vs. negative stimuli and approach vs. avoidance is goal-directed rather than stimulus-driven. For another example, Parkinson (2001) showed that the aggressive act of car honking is sensitive to the expected communicative effect of this behavior. Future research is needed, however, in which the goal-directed processes are shown to determine the initial action tendency (cf. the one-step scenario) rather than being limited to an interventionist role (i.e., the two-step scenario).

## Implications of the Alternative Model for Behavior Change

The alternative view offered here yields important implications for behavior change in clinical practice and society. Many existing dual process models postulate that stimulus-driven processes dominate goal-directed ones because they are more automatic. This implies that therapies designed to regulate suboptimal behavior by trying harder to resist them are doomed to fail because they are in fact attempts to control the uncontrollable. In an effort to escape this gloomy fate, some theorists have proposed not to try harder but to install better habits (cf. approach-avoidance training; e.g., Wiers, Rinck, Kordts, Houben, & Strack, 2010), to re-appraise the stimulus so that it can activate a different [S-R] link, or to automatize self-control (cf. Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). However, if it would turn out that stimulus-driven processes play a minimal role in suboptimal behavior, installing new habits may not be the most promising therapy. Consistent with this idea, approach-avoidance training that is directed at changing habits seems to have mostly short-term effects (Asnaani, Rinck, Becker, Hofmann, 2014; Krypotos, Arnaudova, Effting, Kindt, & Beckers, 2015). But

even cases that do show long-term effects (see Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011) need not be explained by the strengthening of a stimulus-driven process. Repeated execution of an action in a certain context may increase the accessibility of the corresponding action option within the action repertoire (Danner, Vries, & Aarts, 2007) so that it is more likely to be chosen via a goal-directed process.

If research in psychopathology would confirm that various disorders are less driven by rusted stimulus-driven processes, but more by—often hidden—goals, this suggest a therapeutic approach in which problems are analyzed with the aim of uncovering and changing values and expectancies. Several existing strategies fit with this approach. One strategy consists in changing *expectancies* by changing the vividness of outcomes, for instance via simulation or imagination training (Papies & Barsalou, 2015). This provides an opening for long-term goals to gain dominance over short-term ones (when the former are made more vivid than the latter; Watkins, 2011). But imagery may also focus on cultivating immediate goals (e.g., the sensory pleasures of food) so that less consumption is needed to satisfy them (Cornil & Chandon, in press). Another strategy consists in changing the *value* of an outcome or goal (e.g., via reappraisal of the goal). This may be easier to achieve for less valued (lower-order) goals than for more valued (higher-order) ones. Thus, the solution suggested here is not to replace higher-order goals, but to figure out less costly action options for reaching them. For instance, a heavy drinker who does not want to give up feeling high, may switch to doing sport in order to get a runner's high. The alternative view also suggests more promising solutions to societal problems such as prejudice. Whereas classic dual process models focus on automatizing the suppression of prejudice and the installation of new habits (e.g., approach training toward minority members), the alternative view bets on strategies such as imagining the positive outcomes of increased contact (Miles & Crisp, 2014).

## Conclusion

We argued that standard dual process models overestimate the role of stimulus-driven processes in action causation and underestimate the role of their goal-directed counterparts. This is because they hold on to the assumption that stimulus-driven processes are the default process and goal-directed processes intervene only under special conditions. In addition, they tend to overlook a series of complexities when doing empirical research. We believe future thinking and research would benefit if the mappings between processes and operating conditions would be relaxed and the complexities were taken into account. Researchers and practitioners should focus less on stimulus-driven processes when trying to understand so-called suboptimal behaviors and switch their attention to subtle ways in which goal-directed processes could guide those behaviors.

References

Adams, C., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology, 33,* 109-121.

Asnaani, A., Rinck, M., Becker, E., & Hofmann, S. G. (2014). The effects of approach–avoidance modification on social anxiety disorder: A pilot study. *Cognitive Therapy and Research, 38,* 226-238.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37,* 407-419.

Baumeister, R. F. (1997). *Evil: Inside human cruelty and violence.* NY: WH Freeman.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science,* 275, 1293-1295.

Berkowitz, L. (1993). *Aggression: Its causes, consequences, and control*. NY: Mcgraw-Hill.

Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on hostile versus instrumental aggression dichotomy?. *Psychological Review, 108,* 273.

Brandtstädter, J., & Rothermund, K. (2002). The life-course dynamics of goal pursuit and goal adjustment: A two-process framework. *Developmental Review, 22,* 117-150.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review, 97,* 19-35.

Colwill, R. M., & Rescorla, R. A. (1985). Postconditioning devaluation of a reinforcer affects instrumental responding. *Journal of Experimental Psychology: Animal Behavior Processes, 11,* 120-132.

Cornil, Y., & Chandon P. (in press). Pleasure as a substitute for size: How multisensory imagery can make people happier with smaller food portions. *Journal of Marketing Research*.

Danner, U. N., Vries, N. K., & Aarts H. (2007). Habit formation and multiple means to goal attainment: Repeated retrieval of target means causes inhibited access to competitors. *Personality and Social Psycholy Bulletin,* 33, 1367-1379.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience,* 8, 1704-1711.

Eder, A. B. (2011). Control of impulsive emotional behaviour through implementation intentions. *Cognition and Emotion*, *25*, 478-489.

Eder, A. B., & Hommel, B. (2013). Anticipatory control of approach and avoidance: An ideomotor approach. *Emotion Review*, *5*, 275-279.

Eder, A., & Rothermund, K. (2013). Emotional action: An ideomotor model. In C. Mohiyeddini, M. Eysenck, & S. Bauer (Eds.), *Psychology of emotions* (pp. 11-38). Happage, NY: Nova Science Publishers.

Eitam, B., & Higgins, E. T. (2010). Motivation in mental accessibility: Relevance of a representation (ROAR) as a new framework. *Social and personality psychology compass*, *4*, 951-967.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*, 223-241.

Frijda, N. H. (2007). *The laws of emotion*. Lawrence Erlbaum Associates Publishers.

Gerber, J., & Wheeler, L. (2009). A meta-analysis of experimental research on rejection. *Perspectives on Psychological Science*, *4*, 468-488.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, 3*8*, 69-119.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*, 1-26.

Hassin, R. R. (2013). Yes it can: On the functional abilities of the human unconscious *Perspectives on Psychological Science, 8*, 195-207.

Heyes, C., & Dickinson, A. (1993). The Intentionality of Animal Action. In M. Davies, & G. W. Humphreys, (Eds.), *Consciousness: Psychological and philosophical essays* (pp. 105-120). Oxford, UK: Blackwell.

Hommel, B. (2000). The prepared reflex: automaticity and control in stimulus-response translation. In S. Monsell, & J. Driver (Eds.), Control of cognitive processes: Attention and performance XVIII (pp. 247–273). Cambridge, MA: MIT Press.

Keren, G., & Schul, Y. (2009). Two is not always better than one a critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533-550.

Kopetz, C., & Orehek, A. E. (2015). When the end justifies the means: Self-defeating behavior as "rational" and "successful" self-regulation. *Current Directions in Psychological Science, 24*, 386-391.

Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science, 10*, 618-638.

Krypotos, A. M., Arnaudova, I., Effting, M., Kindt, M., & Beckers, T. (2015). Effects of approach-avoidance training on the extinction and return of fear responses. *PloS one, 10*(7), e0131581.

Krieglmeyer, R., De Houwer, J., & Deutsch, R. (2013). On the nature of automatically triggered approach–avoidance behavior. *Emotion Review, 5*, 280-284.

Lang, P. J. (1994). The motivational organization of emotion: Affect-reflex connections. In S. H. M. Van Goozen, N. E. Van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 61-92). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lee, S. W., Shimojo, S., & O'Doherty J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron,* 81, 687-99.

Leigh, B. C. (1989). In search of the Seven Dwarves: Issues of measurement and meaning in alcohol expectancy research. *Psychological Bulletin*, *105*, 361-373.

Matsumoto, D., & Ekman, P. (2009). Basic emotions. *Oxford companion to affective sciences,* 69-72.

Miles, E., & Crisp, R. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes and Intergroup Relations, 17,* 3-26.

Moors, A. (2014a). Examining the mapping problem in dual process models. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 20-34). NY: Guilford Press.

Moors, A. (2014b). Flavors of appraisal theories of emotion. *Emotion Review, 4,* 303-307.

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

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132,* 297-326.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77,* 167-184.

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Science, 18,* 488-96.

Papies, E. K., & Barsalou, L. W. (2015). Grounding desire and motivated behavior: A theoretical framework and empirical evidence. In W. Hofmann & L. F. Nordgren (Eds.), *The psychology of desire*. NY: Guilford.

Parkinson, B. (2001). Anger on and off the road. *British Journal of Psychology, 92,* 507–526.

Parkinson, B. (2008). Emotions in direct and remote social interaction: Getting through the spaces between us. Computers in Human Behavior, 24, 1510-1529.

Saunders, B., Milyavskaya, M., & Inzlicht, M. (2015). Variation in cognitive control as emotion regulation. *Psychological Inquiry*, *26*, 108-115.

Scarantino, A. (in press). Do emotions cause actions, and if so how? *Emotion Review*.

Sheeran, P., Webb, T. L., & Gollwitzer, P. M. (2005). The interplay between goal intentions and implementation intentions. *Personality and Social Psychology Bulletin*, *31*, 87-98.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality & Social Psychology Review, 8*, 220-247.

Tamir, M. (2009). What do people want to feel and why? Pleasure and utility in emotion regulation. *Current Directions in Psychological Science, 18*, 101–105.

Tapia, E., Breitmeyer, B. G., & Shooner, C. R. (2010). Role of task-directed attention in nonconscious and conscious response priming by form and color. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 74-87.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology, 33*, 1313–1330.

Watkins, E. (2011). Dysregulation in level of goal and action identification across psychological disorders. *Clinical Psychology Review*, *31*, 260-278.

Webb, T. L., Schweiger Gallo, I., Miles, E., Gollwitzer, P. M., & Sheeran, P. (2012). Effective regulation of affect: An action control perspective on emotion regulation. *European Review of Social Psychology, 23*, 143-186.

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.

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, *22*, 490-497.

Webb, T. L., Schweiger Gallo, I., Miles, E., Gollwitzer, P. M., & Sheeran, P. (2012). Effective regulation of affect: An action control perspective on emotion regulation. *European Review of Social Psychology, 23*(1), 143-186.

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review, 114*, 843-863.

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology, 67*, 289-314.