Emotions and practical rationality: A comparison of reactive and instrumental theories

Agnes Moors, KU Leuven

Abstract The chapter contrasts two classes of mechanistic emotion theories—reactive and instrumental ones—with a particular focus on how they account for the practical irrationality typically ascribed to emotions. After demarcating practical rationality from theoretical rationality, I distinguish between rationality in the process-sense and the output-sense as well as between direct and indirect paths through which emotions are supposed to influence behavior. Then I describe the mechanisms for emotion causation put forward by the two classes of theories: a stimulus-driven mechanism in reactive theories and a goal-directed mechanism in instrumental theories. Finally, I compare the account of the irrationality of emotions provided by reactive and instrumental theories. I do this separately for the direct and indirect pathways. Regarding the direct pathway, the stimulus-driven mechanism in reactive theories is irrational in the process-sense, but can on occasion yield a rational output. The goal-directed mechanism in instrumental theories is rational in the process-sense, but can still produce an irrational output. Regarding the indirect pathway, reactive theories consider the following biases: emotions as a switch from goal-directed to stimulus-driven processes, incidental emotions, and integral emotions. Instrumental theories proposes an alternative goal-directed explanation for each of these presumed biases that is more parsimonious because it does not require postulating the moderating or mediating role of emotions.

Emotion theories often develop according to an idealized path in which they start from a working definition in which the set of emotions is provisionally demarcated from other sets. A common way to do this is to list a number of typical and apparent properties of the phenomena called emotions. After that, theories come up with constitutive and causal-mechanistic explanations of emotions that should ideally account for these properties. These explanations are then tested in empirical research. Finally, these explanations deliver criteria for the demarcation of the set of emotions from other sets in a scientific definition.

Examples of typical and apparent properties in need of explanation are that emotions have a mental aspect and a bodily aspect. The mental aspect of emotions refers to the fact that they have an object or that they are about something. This is called their Intentional[1] aspect. A commonly accepted distinction in philosophy is that between the particular object and the formal object of emotions (de Sousa 1987; Kenny 1963). For instance, if Tinny is afraid of Tony, the particular object of her emotion is Tony and the formal object of her emotion is danger. If Tony is angry at Tinny for being rude, the particular object is Tinny's being rude

[1] I indicate this meaning of the term Intentionality with a capital 'I' following Searle (1983) to mark the distinction with the ordinary use of intentionality with a lower case 'i'.

and the formal object is offense. If Tony feels guilty about having offended Tinny, the particular object is the fact that he offended Tinny and the formal object is that he transgressed a norm. Emotions are not only mental, but also consciously experienced or felt. This means that in addition to an Intentional aspect, emotions also have a phenomenal aspect. There is something it is like to have an emotion.

The bodily aspect of emotions refers to the fact that they are accompanied by physiological responses as well as by expressive and musculoskeletal behavior. If overt behavior is absent, they are still characterized by an action tendency or urge to act. Emotions are further characterized by a certain 'heat', which means that they have an intense quantity and a positive or negative quality. Furthermore, they are elicited automatically, in the sense that they arise fast, unintentionally, and without requiring a lot of attention, and that they are difficult to counteract. A further property is that they have control precedence, which means that they take priority over mundane goal striving. Finally, emotions have been characterized as possessing an irrational flavor in the sense that they have the potential to bias our thoughts and lead to maladaptive behavior. Although each of these properties can be defined independent of the others, authors often assume relations among them. For example, several philosophers make a strong connection between the felt aspect of emotions and their bodily aspect (e.g., Deonna and Teroni 2012). For another example, many authors assume strong connections between the intense nature of emotions, their automaticity, their control precedence, and their irrational flavor.

This chapter zooms in on the last property in the list: irrationality. While laypeople often highlight the irrational aspect of emotions, emotion theorists seem to agree that emotions can be either rational or irrational. The aim of this chapter is to contrast two classes of mechanistic theories of emotions with a particular focus on how they account for the presumed ir/rationality of emotions. The chapter is structured as follows. In the first section, I break up the question of the relation between emotions and ir/rationality into different parts with the help of a few commonly drawn distinctions such as that between theoretical vs. practical rationality and rationality in the process-sense vs. the output sense. Regarding the practical rationality of emotions, a further distinction can be made between a direct vs. indirect pathway through which emotions can influence behavior. In the second section, I describe the two classes of mechanistic theories of emotions in broad strokes. In particular, I contrast reactive emotion theories with instrumental theories. In the third and final section, I connect the dots by detailing the accounts that both classes of theories have offered for the practical irrationality of emotions.

1. Ir/rationality of emotions

The ir/rationality of an entity is measured relative to certain standards. A common distinction is that between theoretical and practical rationality. In theoretical rationality, an entity is judged as rational if it is *accurate* in representing the external world. This is typically judged for cognitive entities such as perceptions and beliefs. For instance, a perception of a tree is accurate if a tree is indeed present, and the belief that the world is round is accurate if the world is indeed round. Cognitive entities have a mind-to-world direction of fit: They are fitting if they fit with the world, that is, if they are accurate. In practical rationality, on the other hand, an entity is judged as rational if it satisfies a person's goals and leads to well-being. This is typically judged for conative entities such as behavior and goals. For instance, a behavior is practically rational if it satisfies a person's goals, and a goal is itself practically rational if it satisfies a person's superordinate goals or ultimate well-being. Conative entities have a world-to-mind direction of fit: They are fitting if the world fits with them, that is, if they are satisfied.

It should be noted upfront that my definition of rational behavior as behavior that satisfies goals can be aligned with the definition in philosophy of rational behavior as behavior that is 'reason responsive' or done for a reason because reasons can easily be cashed out in terms of goals. If Tinny mows the lawn with the *goal* to impress Tony, the *reason* for her to mow the lawn is to impress Tony. It should nevertheless be emphasized that goals are understood here in a thin sense as representations with dynamic features (e.g., they tend to lead to behavior, their activation grows over time, and they persist in the face of obstacles) and that there is no requirement whatsoever that they are conscious or the result of a deliberate reflective process (contra Wiegman 2020).

Rational behaviors have been contrasted with irrational and arational behaviors (Hursthouse 1991). Irrational behaviors are ones that go against goal satisfaction (i.e., are done for the wrong reasons) whereas arational behaviors are ones that are useless or irrelevant for goals (i.e., are done for no reasons at all). In fact, it can be argued that arational behaviors are still irrational because behavior without any benefits still has costs for the goal to save energy (Moors in press).

Another distinction worth underlining is that between rationality in the output-sense and rationality in the process-sense (Elster 2010). Rationality can be judged not only for the perceptions, beliefs, behaviors, and goals, which can be seen as the outputs of certain processes but also for the processes themselves. A process can be rational but still lead to irrational output and vice versa. For instance, a logical reasoning process can still lead to an

inaccurate belief and an illogical reasoning process to an accurate belief. Likewise, a process designed to satisfy goals can still lead to dissatisfaction and a process that is not so designed to satisfaction. Dissociations between the rationality of processes and their output can occur because the output of a process is not only determined by the process itself but also by its input as well as by other processes that intervene and modify the output of the process.

Applying these notions to emotions, some authors have argued that emotions are cognitive entities like perceptions and beliefs (with a world-to-mind direction of fit), others that they are conative entities like behaviors and goals (with a mind-to-world direction of fit), and still others that they have cognitive as well as conative aspects (with a double direction of fit).

Authors who take emotions to have a cognitive aspect have proposed that they are theoretically rational or accurate if their particular object instantiates the formal object of that emotion. For instance, Tinny's fear is rational if the particular object of her fear (e.g., a wild animal) instantiates the formal object of the emotion fear (i.e., danger). Thus, Tinny's fear of a snake in the wild is rational because snakes in the wild are dangerous whereas her fear of a snake in the zoo is irrational because snakes behind glass are harmless.

Authors who take emotions to have a conative aspect have proposed that emotions are practically ir/rational or mal/adaptive if they push the person towards behavior that dis/satisfies her goals. Thus, the practical rationality of emotions depends on the influence that they have on behavior. Traditionally, it is assumed that emotions can influence behavior via a direct and an indirect pathway.

(1) The direct pathway rests on the idea that each emotion is characterized by an action tendency, which may translate in overt behavior. This behavior is called emotional behavior. Thus fear may lead to flight, anger to fight, guilt to social reparation, sadness to resignation, and so on. If the direct pathway is considered, the practical rationality of emotions seems to partly rely on their theoretical rationality. If fear is accurate, it is more likely to satisfy the person's goals than when it is inaccurate. Thus, Tinny's fear of a snake in the wild is accurate so that fleeing from it satisfies her goal for survival. Her fear of a snake in the zoo, on the other hand, is inaccurate. Her fleeing from it does nothing to satisfy her goal for survival but only comes with the cost that she comes across as silly.

But the relation between theoretical and practical rationality is not one-to-one. Emotions that are accurate can still be maladaptive and emotions that are inaccurate can still be adaptive. As an example of *accurate but maladaptive* emotions, consider the phenomenon of costly aggression, where anger leads to the tendency to hurt others even if it comes at a

personal cost (Pillutla and Murnighan 1996; Sanfey et al. 2003). Another set of examples are emotions leading to arational actions, such as when a person strokes someone's hair out of love, jumps and dances out of joy, slams with the door out of anger, sleeps with the picture of a lost love under her pillow out of grief, and hides her face in the dark out of shame (examples adapted from Hursthouse 1991). These behaviors can all be considered accurate (in the sense that their particular objects instantiate the formal objects of their respective emotions) while still being maladaptive (in the sense that they seem useless while using up some energy).

As examples of *inaccurate but adaptive* emotions, consider cases of instrumental emotions such as a person works up her anger to get what she wants (Greenspan 2004) and a person who swallows her tears to discourage a perpetrator from hurting her even more (Salmela 2008).

(2) The indirect pathway by which emotions can influence behavior rests on the idea that next to emotional behavior, there is also non-emotional behavior, which is caused by non-emotional processes, and that emotions can have a biasing influence on these non-emotional processes. To have an idea what these non-emotional processes of behavior causation are, a brief detour to the behavior domain is warranted. This domain is governed by dual-process theories that distinguish between stimulus-driven and goal-directed processes of behavior causation (Moors, Boddez, and De Houwer 2017). In a stimulus-driven process, the mere presence of a stimulus activates the representation of a stimulus-response association, which in turn translates into overt behavior (S→[S–R]→R). Stimulus-response associations can be innate but they can also be learned in which case they are called habits. In a goal-directed process, a selection is made between different action options based on a weighing of their expected utilities in light of the current stimulus. The expected utility of one action option is the product of the values of the outcomes of the action and the expectancy that the action will indeed lead to these outcomes ($S \rightarrow [S{:}R_1{-}O^v, R_2{-}O^v] \rightarrow [R_1] \rightarrow R_1$).

Dual-process theories come in two versions depending on the way in which they construe the interplay between the two processes. The first version defends a *default-interventionist* architecture (Evans & Stanovich, 2013; Strack & Deutsch, 2004; Wood & Rünger, 2016). According to this version, stimulus-driven processes are simple, in the sense that they trigger an action tendency without considering the outcomes of the corresponding behavior. Their simplicity comes with the benefit that they are automatic but the cost that they are rigid. That they are automatic means that they can operate when there is little opportunity, capacity, and/or motivation to engage in them. That they are rigid means that they cannot update when changes in the actual outcomes of behavior occur. Goal-directed processes, on

the other hand, are seen as more complex, which comes with the benefit that they are more flexible but also the cost that they are nonautomatic. In this architecture, stimulus-driven processes are the default process because they are automatic. Because of their rigidity, however, stimulus-driven processes are also prone to produce maladaptive behavior. This is where goal-directed processes come in. Their role is to intervene when stimulus-driven processes threaten to lead to maladaptive behavior and to correct the course of action. However, because of the nonautomatic nature of goal-directed processes, they can only operate when there is ample opportunity, capacity, and motivation. To illustrate, the sight of the bowl of nuts is supposed to trigger the tendency to eat from it even when one is not hungry. If the person is absorbed in a conversation and therefore distracted, grabbing nuts takes place without hindrance. If she is attentive and motivated to go against it, the nut-grabbing may be suppressed.

The second version of dual-process theories defends a *parallel-competitive* architecture (Moors et al. 2017) According to this version, both stimulus-driven and goal-directed processes can be automatic, which is why they often operate in parallel and compete with each other. In addition, this version also postulates that the competition will most often be won by the goal-directed process. According to this version then, goal-directed processes are the default determinant of behavior whereas stimulus-driven processes are the exception. It may further be noted that the competition between stimulus-driven and goal-directed processes often takes place below the surface of consciousness. Here, self-regulation conflicts that one does experience will often reflect a conflict between two goal-directed processes. Grabbing nuts may be instrumental for the goal to have a pleasant taste or the goal to keep one's hand busy while staying clear from the nuts may be instrumental for the goal to avoid a full stomach or the goal to put on more weight.

Now that these distinctions are in place, I will discuss two classes of emotion theories, and I will zero in on the way in which they account for the practical irrationality that is often attributed to emotions.

2. Two classes of emotion theories

Emotion theories come in many shapes and shades. If we care to look beyond the surface, however, an important fault line can be drawn between these theories.

2.1. Reactive theories

A first class of theories, which I will call reactive theories, build on the mechanism for emotion causation proposed by evolutionary theories. This mechanism takes the form of a stimulus-driven process. Incoming stimuli activate an innate association between a stimulus

representation and a response representation or action tendency. This association, also known as an affect program, is supposed to be a remnant of the behavioral solutions that our evolutionary ancestors selected to recurrent challenges and opportunities in their environment. The association has been framed by some theories in concrete terms, as in 'snake-flee', and by others in more abstract terms, as in 'danger-defense'. In 'concrete' reactive theories, a snake is first perceived in concrete terms (as a snake, or as a slippery creature) and this in turn activates a fixed link between the representation of the snake and the tendency to flee. Once the affect program is activated, it subsequently triggers physiological responses preparing the organism to flee, characteristic facial and vocal expressions, actual fleeing behavior, and feelings of fear. In 'abstract' reactive theories, the snake is first evaluated as dangerous and this in turn activates a fixed link between the representation of danger and the tendency to seek safety. In the latter version, the affect program is preceded by a stimulus evaluation process also known as an appraisal process.

In addition to an innate stimulus-driven process, which accounts for the pure emotion or 'hard liquor', reactive theories still allow goal-directed processes involved in emotion regulation and planning to modify or refine the course of events. For instance, if Tinny has the initial tendency to flee from a snake in the zoo, this may be incongruent with her goal to make a mentally sane impression, allowing her to suppress her goal to flee. If Tinny encounters a snake in the wild, and there is no reason to suppress her fleeing from the snake, she still has to decide whether she will jump over the fence on her right or hide behind the tree on her left. Here again, a goal-directed process is involved, but this time in the form of a planning process.

Importantly, reactive theories endorse a *default-interventionist* perspective on the interplay between stimulus-driven and goal-directed processes. The stimulus-driven processes that undergird the pure emotion are the default, whereas the goal-directed processes involved in regulation and planning can only intervene or supplement this process. This division of labor is rooted in the deep-seated conviction that stimulus-driven processes are automatic whereas goal-directed processes are slow and laborious. This is why the initial action tendency is assumed to be caused by a stimulus-driven process and why goal-directed processes are relegated to a later point in time and in a regulatory or implementational role.

2.2. Instrumental theories

A second class of emotion theories, which I will call instrumental theories, propose that goal-directed processes are not just involved in emotion regulation and planning but also in emotions themselves. This class of theories has fewer members, including early and later

reinforcement learning theories of emotions (early: Rolls, 2005; later: Broekens et al. 2015) and my own goal-directed theory of emotions (Moors 2017a, 2017b; Moors et al. 2017). The latter theory is a dual-process model with a *parallel-competitive* architecture. The theory assumes that stimulus-driven and goal-directed processes can both be automatic and therefore often operate in parallel and compete with each other. In addition, the theory postulates that goal-directed processes are stronger than most stimulus-driven ones and will therefore often win the competition. The net result is that most of our behavior, whether emotional or non-emotional, will be caused by goal-directed processes. This means that actions are selected based on their expected utilities, that is, based on a foreshadowing of the values of the outcomes of these actions and the likelihood that these outcomes will occur.

The goal-directed theory has elaborated on this proposal by embedding this goal-directed weighing process within a broader goal-directed cycle. The cycle starts with the comparison between a stimulus and a first goal. In case of a discrepancy, a second goal is activated aimed at reducing the discrepancy. The second goal can be achieved via three broad strategies: (a) by choosing a behavior (i.e., assimilation), (b) by swapping the first goal by a different one (i.e., accommodation), or (c) by reinterpreting the stimulus so that it is less discrepant with the first goal (i.e., immunization). Both the selection of strategies of assimilation, accommodation, and immunization, as well as the selection of a specific action option in the case of assimilation are assumed to be determined by the weighing up of the expected utilities of these strategies and action options. To illustrate, Tinny wants to go for a ride with her bike but notices that the tire pressure is very low. To reduce this discrepancy, she can choose to engage in behavior (i.e., assimilation), in which case she still has to decide whether she will merely pump up her tire or scan the tire for holes and repair them. She can also decide to go for a walk instead of a bike ride (i.e., accommodation). Finally, she can judge that her tire is not that flat and go for a bike ride after all (i.e., immunization).

The theory assumes that the goal-directed cycle involved in episodes of mundane goal pursuit is the same as those involved in emotional episodes. To illustrate, Tony wants to be respected by his boss, but the boss insults him during a meeting. To reduce this discrepancy, Tony can choose to engage in behavior (i.e., assimilation), in which case he still has to decide whether he will fight back or flee from the situation. He can also decide to give up his goal to be respected by his current boss and apply for a different job. Finally, he can tell himself that his boss merely tried to play the ball not the man, and hence that his goal to be respected was in fact not violated.

As mentioned, the theory holds that the mechanisms involved in emotional episodes are exactly the same as those involved in non-emotional episodes. The difference between both types of episodes lies in the value of the goals that are at stake, but this is merely a matter of degree. Thus, the goal to ride a bike may have a lower value than the goal to be respected by one's boss. This entails that the stimulus-goal discrepancy in the bike example is smaller and perhaps less urgent than that in the boss example, which in turn also translates into a weaker action tendency. When Tinny discovers the low pressure of her bike's tires, she may select an action option that induces a weak action tendency that does not produce any palpable physiological responses. When Tony detects the insult, on the other hand, he may select an action option that induces a vigorous action tendency that produces blatant physiological responses.

The lack of a deep, mechanistic, distinction between emotional and non-emotional episodes has led me to conclude that the set of emotions may not qualify as an adequate scientific set. Whenever I speak in the name of this theory, the term emotion always refers to the folk concept of emotion—the phenomenon that people regard as emotion—not to a scientific category.

3. Connecting the dots

In this section, I examine how reactive and instrumental emotion theories account for the practical irrationality often attributed to emotions. In particular I compare reactive theories with my own goal-directed theory. I revisit the direct and indirect pathways by which emotions are supposed to influence behavior and I examine how the two classes of theories make sense of the phenomena that fall under these two pathways.

3.1. Direct pathway

In the direct pathway, each emotion is supposed to activate a behavior that satisfies the action tendency characteristic of the emotion. Anger activates aggression, which satisfies the tendency to fight, whereas fear activates fleeing, which satisfies the tendency to seek safety, and so on. Of course, the direct pathway from emotion to behavior is only meaningful for theories that take emotions to be proper scientific entities. This is the case for reactive theories but not for the goal-directed theory. After presenting the account of reactive theories, I clarify how the goal-directed theory manages to explain these phenomena without invoking emotions.

a. Reactive theories

In reactive theories, the backbone of each emotion is a fixed, innate link between a stimulus representation and a response representation or action tendency. Stimulus-driven processes

are maladaptive in the process-sense because they do not take into account the outcomes of behavior. This being said, stimulus-driven processes can still be adaptive or maladaptive in the output-sense. This means that the behavioral solutions to the challenges of our evolutionary ancestors may turn out to be beneficial or costly in the current context of the organism. For instance, if Tinny's fear makes her jump back from a snake in the wild, it is adaptive. If it makes her jump back from a snake in the zoo, making her the subject of ridicule, it is maladaptive. Note that in these examples, the practical rationality in the output-sense hinges on the theoretical rationality of the emotion. Tinny's fear of the snake in the wild is *adaptive* because it is *accurate* and her fear of the snake in the zoo is *maladaptive* because it is *inaccurate*.

In cases in which the theoretical and practical rationality come apart, emotions may also be rational or irrational in the output sense. Let us consider the case of costly aggression. If Tony's anger at his boss's insult pushes him into an aggressive outburst and he gets fired afterwards, his anger turns out to be maladaptive. If, after the initial shock, the boss starts treating him with more respect, his anger may still prove to adaptive after all. In these examples, Tony's anger about being insulted is *accurate* regardless of how his boss reacts and hence regardless of whether this anger turns out to be *maladaptive or adaptive*.

A similar story can be told for arational actions, which are thought to fulfill the action tendencies characteristic of emotions. Some arational actions are said to be fully driven by emotions. Examples are stroking someone's hair out of love, which fulfills love's action tendency to seek proximity, jumping and dancing out of joy, which fulfills joy's action tendency to broaden and build, and hiding one's face in the dark, which fulfills shame's action tendency to hide oneself. Other arational actions are said to be partly driven by emotions but partly also by regulation or planning. Examples are slamming the door out of anger and talking to a lost love's picture out of grief. Slamming the door partly fulfills the action tendency of anger, which is to attack, but it is partly also regulated to prevent retaliation by displacing it to a lifeless object. Talking to a lost love's picture partly fulfills the action tendency of grief to seek proximity to the lost person, but which for lack of availability of this person, implements this tendency by displacing it to a closely related object. In all these cases, the emotions may be accurate regardless of whether the consequent behaviors turn out to be adaptive or maladaptive. If Tony slams the door once he's returned home, and the aggression resulting from his anger comes with a slight energetic cost, his anger qualifies as slightly maladaptive. If, on the other hand, his aggression helps reduce part of his excess adrenaline, his anger counts as slightly adaptive.

But what about *inaccurate but adaptive* emotions such as those in which people fake or suppress an emotion to reach a certain goal? In the boss example, Tony may have worked up his anger with the intention to earn his boss's respect in the long run or he may have successfully prevented becoming angry to prevent the slightest risk of retaliation. Reactive theories do account for cases like these—in which the person does envisage the outcomes of her behavior—by calling on the goal-directed process of emotion regulation. Emotion regulation is not restricted to the suppression of an ongoing emotion's outward manifestation, called consequent down-regulation. It also covers consequent up-regulation, such as in the example of Tony working up his anger, and antecedent down- and up-ward regulation, as in examples of Tony nipping his anger in the bud or faking to be happy. It is important to underline, however, that reactive theories make a strict separation between the goal-directed process involved in emotion regulation and the stimulus-driven process involved in the emotion itself.

b. Instrumental theories

In the goal-directed theory, the mechanism involved in so-called emotional episodes is as goal-directed as that involved in episodes of mundane goal pursuit. Given that goal-directed processes are the textbook example of a practically rational process, we can infer that this theory takes so-called emotions to be practically rational in the process-sense. This begs the question, however, of how this theory can account for the apparent practical irrationality or maladaptiveness of many of our so-called emotions. The theory proposes three possible answers to this question. A first reply is that there are cases of adaptive emotions that appear maladaptive to the outside world or even to the emoter. If Tony's goals of being respected is currently more dear to him than his goal of staying on the job, then bursting out into aggression may in fact be the best action option he could choose at that time.

A second reply is that people may engage in a goal-directed process but make mistakes along the way. They may start off with an over- or underestimation of the value of the first goal in the cycle. This is because the first goal in the cycle is itself the result of a selection process in which its expected utility for reaching a superordinate was weighted relative to that of other goals. People may also fail to register that an encountered stimulus forms a discrepancy with the first goal. Even if a discrepancy is registered, people may still choose the wrong strategy (e.g., immunize when assimilation is better) or choose the wrong action option (e.g., fight when fleeing is better) because the subjective expected utilities of strategies and action options deviate from their objective expected utilities. Rationality is never Olympian but bounded by limited access to information, limited processing capacities,

and limited motivation for processing the information that is accessible (Simon 2000). Finally, even if the best action option does in fact get chosen, things may go wrong during implementation of this action. Altogether, this second reply can be read as saying that emotions are practically rational in the process-sense, but that because this rationality is bounded, they may turn our to be practically irrational in the output-sense.

A third reply is that even if the person would flawlessly traverse the goal-directed cycle, the chosen action may not yield the expected outcome because unexpected obstacles may throw a spanner in the works. Tony's selection of aggressive behavior may objectively be the best way to reduce the discrepancy with his cherished goal of being respected by his boss, but it may never come to this if his boss gets struck by lightning the next day. This third reply can be read as saying that even if practical rationality in the process-sense would be Olympian, there would still be a way to account for cases of practically irrational emotions in the output-sense.

The above-described replies can account for cases of irrational emotions, but what about cases of arational emotions? The goal-directed theory explains all forms of arational actions in terms of goal-directed processes without having to call on the deus-ex-machina solution that these are caused by emotions or the stimulus-driven processes underlying them. This means that for each of these actions, it should be plausible to assume that it is selected because it has a high expected utility to reach a certain goal. Stroking someone's hair may be selected as the best way to reach the goal for proximity and this goal for proximity may itself be selected against other goals instead of being driven by an emotion of love. Likewise, jumping and dancing after goal achievement may be instrumental for the goal to open up to opportunities for the fulfilment of other goals. Alternatively, it may be instrumental for the goal to communicate one's success to others. This explanation is not compromised by cases in which people jump and dance after hearing good news when they are alone. This is because people may still communicate to an audience in their head (Fridlund et al. 1990). In fact, the goal to communicate to a real or imaginary audience can account for each of the remaining cases of arational actions such as covering one's face in the dark after having done something wrong, slamming the door after having been offended, and talking to a lost love's picture. In sum, all of these so-called arational actions may be goal-directed so that there is no need to invoke emotions as explanatory entities.

3.2. Indirect pathway

a. Reactive theories

In the indirect pathway, emotions are supposed to bias the mental processes involved in the causation of non-emotional behavior. To reiterate, dual-process models in the behavior domain propose that non-emotional behavior can be caused by a stimulus-driven or a goal-directed process. In the literature on emotions and rationality, which is dominated by reactive emotion theories, several types of biases have been listed.

One bias is that emotions act as a switch between stimulus-driven and goal-directed processes involved in non-emotional behavior. Based on the assumption that stimulus-driven processes are automatic and goal-directed processes nonautomatic, it is assumed that when a person's processing capacity is reduced, she can no longer engage in goal-directed processes so that her behavior will be dictated by stimulus-driven ones. Emotions are considered to be one source of reduced processing capacity. If Tinny is consumed by fear or distress, she mindlessly follows her ingrained habits (Schwabe and Wolf 2011).

Another set of biases is that emotions influence the representations involved in each of the non-emotional processes of behavior causation. Research has typically restricted its focus on the representations involved in the goal-directed processes governing non-emotional decision-making. Here, a distinction is commonly made between incidental emotions and integral ones (Lerner et al. 2015).

An *incidental* emotion is one that influences a decision while being unrelated to this decision. Here, emotions often take on the role of a moderator of the relation between a decision-making situation and the resulting decision. For instance, Tinny just experienced fear from a car that drifted away from its lane and is now faced with the decision to switch to the fast lane or to remain in the slow lane. Her fear keeps her in the slow lane. The explanation of reactive emotion theories is that emotions—or the stimulus evaluations and/or action tendencies that are part of them—get carried over to the subsequent decision process.

An *integral* emotion is one that influences a decision while being related to this decision. Here, emotions often take the role of a mediator between the decision-making situation and the resulting decision. For instance, Tinny fears that switching to the fast lane will create the risk of a collision with another car. As this example shows, integral emotions are often anticipatory in that they result from anticipating the future outcome of a decision. This outcome can be material or emotional in nature. Thus, Tinny's fear may be about a car crash or about the regret she will experience after her car has crashed. Integral, anticipatory emotions must be differentiated from anticipated emotions. The former are experienced at the time of the decision whereas the latter are not and are therefore better called anticipations of

emotions. An example of an anticipation of an emotion is when Tinny expects or anticipates that she will regret a car crash without currently experiencing fear about this prospect.

b. Instrumental theories

The goal-directed theory proposes that the phenomena that others call emotions are nothing but goal-directed decision processes or parts thereof. This section revisits the biases that fall under the heading of the indirect pathway by which emotions are supposed to influence behavior and considers the alternative explanations that are offered by the goal-directed theory.

Let us start with the first bias, according to which emotions are supposed to act as a switch between stimulus-driven and goal-directed processes. The observation that requires an explanation is that people who are consumed by emotions are more likely to engage in suboptimal behavior. The alternative explanation offered by the goal-directed theory is that the person is consumed by a goal-directed process that is at the service of a highly valued goal. If a person's attention is concentrated on reducing one goal-discrepancy, there will be less capacity or motivation to spend on reducing another one.

The second bias to consider is the influence of *incidental* emotions on decision making. Let us revert to the example of the lingering fear Tinny experienced after just having avoided a car accident and which influenced her decision to keep driving in the slow lane for a while. Instead of accepting that Tinny was indeed overcome by the emotion of fear and that this fear (or the evaluation of danger and/or the avoidant action tendency characteristic of this fear) carried over to a subsequent decision process, the goal-directed theory proposes that the influence of incidental emotions on decision making can be reframed as cases in which a first decision process influences a second one. The nearly missed discrepancy between the stimulus and the goal for safety registered in the first decision process may have raised the salience and/or value of this goal so that the goal for safety also dominated the second decision process. Another option is that the avoidant action tendency that was selected in the first decision process was salient in the person's action repertoire and was therefore easily selected again in the second decision process (see Moors and Fischer 2019).

The third bias to discuss is the influence of *integral* emotions on decision making. Take again Tinny's fear that switching to the fast lane will lead to a car crash and/or to regret. Instead of accepting that the prospect of a car crash or regret guides Tinny's decisions because they are the particular object of Tinny's fear, these outcomes simply constitute anticipated discrepancies with the goal for safety and/or the goal to feel good. No calling upon a specific emotion as a mediator is required. Note, moreover, that the regret in this example need not

exist in reality, but only in Tinny's mind. Anything can become the content of a goal including the goal to feel good or not feel regret. As long as a person believes that regret is an entity that exists and is to be avoided, it can have the power to guide her behavior.

4. Conclusion

The majority of emotion theories belong to the class of reactive theories. These theories are marked by the explicit or implicit assumption that emotions are undergirded by stimulus-driven processes that are rigid, even if they can be altered by goal-directed regulation processes or refined by goal-directed planning processes. It is the rigidity of stimulus-driven processes that makes emotions practically irrational in the process-sense. Depending on the context in which emotions are elicited, the behavior they produce can turn out to be adaptive or maladaptive. In addition to this direct pathway by which emotions influence behavior, reactive theories also describe an indirect pathway by which they do so. Here, emotions are called upon as entities that bias goal-directed processes involved in mundane decision making, either in the role of a switch, a mediator, or a moderator.

The goal-directed emotion theory, which is one member of the class of instrumental theories, proposes that the phenomena that people call emotions rely on goal-directed processes for action selection that do not fundamentally differ from those involved in mundane goal-pursuit. They may differ in the value of the goals involved and hence the intensity and/or urgency of the action tendencies elicited, but this is only a matter of degree. The implications of this view are that emotions do not qualify as scientific entities and that they are rational in the process-sense. If our experience tells us otherwise, this is simply because goal-directed action selection largely operates below the surface of consciousness and is often at the service of hidden goals. This said, the theory can still account for 'emotions' that are genuinely irrational in the output sense. It does so by arguing that maladaptive behavior is caused by sand in the wheels of the goal-directed process. While the previous relates to the so-called direct pathway from emotion to behavior, the goal-directed theory also provides an alternative account for the biases taken to belong to the indirect pathway. If the proposal is taken at heart that emotions are nothing but episodes of decision-making, any moderating or mediating influence of emotions on decision-making can be reframed as the influence of one decision process (or factors therein) on another decision process. The goal-directed theory offers a parsimonious alternative against mainstream reactive theories in that it dispenses with emotion as an explanatory construct. This not only contributes to the demystification of cases of irrational behavior, but also provides clear handles for behavior

change. If maladaptive behavior is caused by inaccuracies in values and expectancies, correcting these inaccuracies provide a way forward.

References

Broekens, Joost, Elmer Jacobs, and Catholijn M. Jonker. "A Reinforcement Learning Model of Joy, Distress, Hope and Fear." *Connection Science* 27, no. 3 (2015): 215-233. https://doi.org/10.1080/09540091.2015.1031081

Deonna, Julien, and Fabrice Teroni. *The Emotions: A philosophical Introduction*. London: Routledge, 2012. https://doi.org/10.4324/9780203721742

de Sousa, Ronald. *The Rationality of Emotion.* Cambridge, MA: MIT Press, 1987. https://doi.org/10.7551/mitpress/5760.001.0001

Elster, Jon. "Emotional Choice and Rational Choice." In *The Oxford handbook of philosophy of emotion*, edited by Peter Goldie, 263-281. Oxford: Oxford University Press, 2010. https://doi.org/10.1093/oxfordhb/9780199235018.003.0012

Evans, Jonathan. St. B. T., and Keith E. Stanovich. "Dual-process Theories of Higher Cognition Advancing the Debate." *Perspectives on Psychological Science*, *8,* no. 3 (2013): 223–241. https://doi.org/10.1177/1745691612460685

Fridlund, Alan J., John P. Sabini, Laura, E. Hedlund, Julie, A. Schaut, Joel I. Shenker, and Matthew J. Knauer. "Audience Effects on Solitary Faces during Imagery: Displaying to the People in your Head." *Journal of Nonverbal Behavior* 14, no. 2 (1990): 113-137. https://doi.org/10.1007/bf01670438

Greenspan, Patricia. "Practical Reasoning and Emotion." In *The Oxford Handbook of Rationality* , edited by Alfred R. Mele and Piers Rwaling, 206-221. Oxford: Oxford University Press, 2004. https://doi.org/10.1093/oxfordhb/9780195145397.003.0011

Hursthouse, Rosalind. "Arational Actions." *The Journal of Philosophy* 88, no. 2 (1991): 57-68. https://doi.org/10.2307/2026906

Kenny, Anthony. *Action, Emotion and Will*. London: Routledge, 2003. https://doi.org/10.4324/9780203711460

Lerner, Jennifer S., Piercarlo Valdesolo, and Karim S. Kassam. "Emotion and Decision Making." *Annual Review of Psychology* 66 (2015): 799-823. https://doi.org/10.1146/annurev-psych-010213-115043

Moors, Agnes. *Demystifying Emotions: A Typology of Theories in Psychology and Philosophy*. Cambridge, UK: Cambridge University Press, in press.

Moors, Agnes. "Integration of Two Skeptical Emotion Theories: Dimensional Appraisal theory and Russell's Psychological Construction Theory." *Psychological Inquiry* 28, no. 1 (2017): 1-19. https://doi.org/10.1080/1047840X.2017.1235900

Moors, Agnes. "The Integrated Theory of Emotional Behavior Follows a Radically Goal-directed Approach." *Psychological Inquiry* 28, no.1 (2017): 68-75. https://doi.org/10.1080/1047840X.2017.1275207

Moors, Agnes, Yannick Boddez, and Jan De Houwer. "The Power of Goal-Directed Processes in the Causation of Emotional and Other Actions." *Emotion Review* 9, no. 4 (2017): 310-318. https://doi.org/10.1177/1754073916669595

Moors, Agnes, and Maja Fischer. "Demystifying the role of emotion in behaviour: Toward a goal-directed account." *Cognition and Emotion* 33, no.1 (2019): 94-100. https://doi.org/10.1080/02699931.2018.1510381

Pillutla, Madan M., and J. Keith Murnighan. "Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers." *Organizational Behavior and Human Decision Processes* 68, no. 3 (1996): 208-224. https://doi.org/10.1006/obhd.1996.0100

Rolls, Edmund T. *Emotion Explained*. Oxford: Oxford University Press, 2005. https://doi.org/10.1093/acprof:oso/9780198570035.001.000

Salmela, Mikko. "How to Evaluate the Factual Basis of Emotional Appraisals?." *Fact and value in emotion* , edited by Louis C. Charland and Peter Zachar, 35-51. Amsterdam: John Benjamins. https://doi-.org/10.1075/ceb.4.03sal

Sanfey, Alan G., James K. Riling, Jessica A. Aronson, Leigh E. Nystrom, and Jonathan D. Cohen. "The Neural Basis of Economic Decision-Making in the Ultimatum Game." *Science* 300, no. 5626 (2003): 1755-1758. https://doi.org/10.1126/science.1082976

Schwabe, Lars, and Oliver T. Wolf. "Stress-Induced Modulation of Instrumental Behavior: From Goal-Directed to Habitual Control of Action." *Behavioural Brain Research* 219, no. 2 (2011): 321-328. https://doi.org/10.1016/j.bbr.2010.12.038

Searle, John R. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, 1983. https://doi.org/10.1017/cbo9781139173452

Simon, Herbert A. "Bounded Rationality in Social Science: Today and Tomorrow." *Mind & Society* 1, no. 1 (2000): 25-39. https://doi.org/10.1007/BF02512227

Strack, Fritz, and Roland Deutsch. “Reflective and Impulsive Determinants of Social Behavior.” *Personality and Social Psychology Review*, *8*, no 3 (2004): 220-247. https://doi.org/10.1207/s15327957pspr0803_1

Wiegman, Isaac. "Emotional Actions Without Goals." *Erkenntnis* (2020): 1-31. https://doi.org./10.1007/s10670-019-00200-8

Wood, Wendy, and Dennis Rünger. “Psychology of Habit.” *Annual Review of*

Psychology, *67* (2016): 289–314. https://doi.org/10.1146/annurev-psych-122414-033417