

Proactive control and agency

René Baston

Article - Version of Record

Suggested Citation:

Baston, R. (2022). Proactive control and agency. *Phenomenology and the Cognitive Sciences*, 23, 43–61.
<https://doi.org/10.1007/s11097-022-09830-w>

Wissen, wo das Wissen ist.



UNIVERSITÄTS- UND
LANDESBIBLIOTHEK
DÜSSELDORF

This version is available at:

URN: <https://nbn-resolving.org/urn:nbn:de:hbz:061-20260721-115330-4>

Terms of Use:

This work is licensed under the Creative Commons Attribution 4.0 International License.

For more information see: <https://creativecommons.org/licenses/by/4.0>



Proactive control and agency

René Baston¹ 

Accepted: 6 June 2022 / Published online: 1 July 2022
© The Author(s) 2022

Abstract

Can agents overcome unconscious psychological influences without being aware of them? Some philosophers and psychologists assume that agents need to be aware of psychological influences to successfully control behavior. The aim of this text is to argue that when agents engage in a proactive control strategy, they can successfully shield their behavior from some unconscious influences. If agents actively check for conflicts between their actions and mental states, they engage in reactive control. For engaging in reactive control, agents need awareness of those mental states which are in conflict with an action. In contrast, if agents are actively maintaining a goal in consciousness, they engage in proactive control. Proactive control does not consist of conflict detection or conflict resolution. I argue that proactive control explains how agents overcome unconscious psychological influences. In doing so, I claim that consciousness is important for engaging in reactive and proactive control.

1 Proactive and reactive control

The neuropsychologist Braver (2012) distinguishes between proactive and reactive control. Both, says Braver, ground the ability to regulate and coordinate thought and behavior to achieve goals. A subject engages in *reactive control* if the goal is only reactivated because a conflict is recognized. For example, an agent has the goal to get some petrol after work, and she is on her way home by car. While she passes the petrol station, she notices a conflict between the current action plan to drive home, and the previous plan to get petrol. She resolves the conflict by effortful overriding the action plan to drive home with the plan to get petrol instead. Thereby, reactive control is a direct response to an actual encountered problem.

✉ René Baston
rene.baston@hhu.de

¹ Heinrich-Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany

In contrast, if an agent engages in *proactive control*, she actively maintains the goal to get petrol in consciousness. Prior to the demanding event, the agent keeps the goal occurrent. Because this occurrent goal-representation has top-down influences on attention, perception and memory, the agent has no conflicting action-plans: As soon as she sees the petrol station, she drives by because she knew from the beginning that she will, first, get petrol and afterwards drive home.

In philosophy, many theorists focus on *reactive control* when they consider the agents' ability to exercise control over their actions. Levy (2014a, b), for instance, claims that control presupposes that an agent recognizes whether an action or a mental state's content is consistent with an agent's personal-level attitudes. For Levy, it is necessary for an agent to be in control that the agent is conscious of the action's moral significance, or that the agent is conscious of the mental states' contents (and influences), which give an action its moral significance. For example, an agent recognizes the moral significance of an action, if she notices in a certain situation that taking the car keys is actually *stealing* the car keys. Only if an agent is aware of these aspects, the agent can check for existing conflicts.

This understanding of control is *reactive* because having control requires *awareness of the conflict* and the ability to solve it. For example, on the morning of May 24, 1987, Kenneth Parks killed his parents-in-law while sleepwalking (Levy 2014b, p. 71). He could not recall what happened that morning, pleaded not guilty, and was found not guilty. Parks was not in control as he could not recognize the conflict between the action-plan to kill his parent-in-law and his plan or policy not to hurt them. While there are different types of potential conflicts, the relevant conflict-types in this text are those in which an agent has conflicting action-related mental states, like a conflict between an association that triggers automatic behavior dispositions and an agent's action-plan.

Reactive control presupposes an agent's self-knowledge of conflicting elements because being aware of reasons, stimuli or actions is necessary to recognize conflicts. In contrast, proactive control does not depend on conflict detection.¹ Agents can engage in proactive control if they actively maintain a goal-representation in consciousness and, thereby, pay more attention to goal-congruent properties of the situation.²

Psychologists widely assume that we are unconscious of various psychological influences on us (Bargh, 2017; Bargh & Morsella, 2008; Custers & Aarts, 2010; Wilson, 2004). For example, the bystander-effect (Darley & Latané, 1968) refers to the phenomenon that a subject's likelihood of helping decreases if passive bystanders are present in a critical situation. According to the study, the participants "reported that they were aware that other people were present, but they felt that this made no difference to their own behavior" (Darley & Latané, 1968, p. 381). However, while participants are unaware of which stimuli influences them, participants are aware of

¹ As it will turn out later, proactive control is impossible without certain types of self-knowledge. For example, agents need to know their own plans or values. Furthermore, agents need to know which situations are potentially relevant for their plans or values. However, proactive control does not presuppose that agents are aware of influencing psychological mechanisms that lead to conflicts with action-plans.

² Possibly, a different type of proactive control is to *anticipate* a conflict between an action and personal-level values. In such a case, whether there is a conscious or unconscious conflict does not matter.

them not helping, although they cannot access some of the explanatory reasons (Sandis, 2012). As they self-monitor their behavior, they are arguably in control of their behavior, because they can align their monitored reactions with their acknowledged values (Mele, 1995).

In contrast, implicit biases lead to *unintentional* discrimination, as biased agents are unaware of their discriminatory behavior. For example, in one experiment that illustrates implicit biases subjects evaluate job applications and unintentionally engage in discriminatory behavior as they assess the job applications from Black (Dovidio & Gaertner, 2000) or Arab-Muslim (Rooth, 2010) applicants worse than those of others just because of the applicants' social groups. As the participants were unaware of their discriminatory behavior, they could not detect a conflict between their shown behavior and their values. Because agents are unaware of the psychological influences from implicit prejudices or implicit stereotypes, they cannot control their behavior in such cases (Holroyd, 2012; Levy, 2014a).

However, as Braver suggested, exercising control goes beyond detecting and resolving conflicts. We only get a full picture of how agents exercise control if ways of proactive control are considered. I think that this full picture solves a problem: although agents are unaware of unconscious psychological influences, like implicit biases, they often successfully exercise control over their behavior. As I will show in Sect. 3, having implicit prejudices or implicit stereotypes does not necessarily translate to discrimination. Additionally, agents who are internally motivated to control their prejudiced reactions show less biased behavior. If, however, control presupposes that agents are aware of conflicting psychological influences, then we cannot account for that.

In this paper, I want to explain the problem and suggest how to solve it. First, the problem:

- a) Agents are *unaware* of implicit biases.
- b) For exercising control, *reactive control* presupposes awareness of implicit biases.
- c) Some agents successfully exercise control over their implicit biases despite (a) and (b).

I shall argue for the following:

- 1. Agents are *unaware* of implicit biases (Sect. 2).
- 2. Some agents can control their behavior, although they have implicit prejudices (Sect. 3).
- 3. Proactive control explains how agents actively shield their behavior from unconscious psychological influences (Sect. 4 and Sect. 5).

In Sect. 2, I will briefly introduce the concept of implicit biases. Here, I will distinguish between different awareness aspects of implicit mental states and their effects. While agents might have awareness of their implicit prejudices' contents, they are unaware of implicit prejudices' effects. In Sect. 3, I argue that the internal motivation to control prejudiced reactions decreases the influence of implicit prejudices and implicit stereotypes. *This is puzzling if agents only exercise reactive control.*

Section 4 will show that if agents have to be aware of conflicting behavioral plans to exercise control, then implicit biases are uncontrollable. In Sect. 5, I will explain how proactive control can account for control while unconscious psychological mechanisms are at play. In particular, I will show that actively maintaining a goal representation in consciousness has top-down effects on attention. These effects, in turn, explain how an agent shield herself from unconscious psychological influences. I will show with empirical data that shifting one's attention to goal-relevant clues excludes certain conscious and unconscious reasons from processing. Furthermore, I explain why proactive control is an aspect of agency, as actively maintaining a goal in consciousness is something that agents *do*. I shall argue that proactive control is an intentional activity performed by agents.

Proactive control, I shall argue, has implications for implicitly biased behavior. An agent can unintentionally discriminate against a social group because of implicit biases which undermine the agent's control. For instance, the study from Rooth (2010) shows such a characteristic case of implicit biases. In the study, participants had to evaluate job applications of Swedish and Arab-Muslim men. The psychologists measured the participants' implicit attitudes with the Implicit Association Test (Greenwald et al., 1998). Those participants with implicit prejudices judged applications from Arab-Muslim men as less good than applications from Swedish men of equal quality. In such a case, the agent unintentionally discriminates against a social group, and the moral significance of the action, namely to discriminate, is inaccessible to the agent.

2 The agent's unawareness of implicit biases

Currently, the notion of 'implicit biases' is discussed in social psychology as well as in philosophy. While there are different possible definitions of implicit biases (Holroyd et al., 2017) the most natural understanding is that implicit biases are the subject's unintended behavioral patterns that contradict the subject's acknowledged goals or values, like egalitarianism.

Characteristically, but not necessarily, implicit attitudes are the causes of implicit biases. In social psychology an attitude is an overall evaluation of an object (Haddock & Maio, 2015). An attitude consists of three components: the affective, the cognitive and the behavioral component. Negative affective attitudes towards a social group are prejudices. Cognitive information about a social group, like 'African-Americans are athletic', builds a stereotype. According to the literature on implicit social cognition, a single person can have a positive *explicit* attitude towards a social group, while having a negative *implicit* attitude towards the same group at the same time (e.g., Banaji 2001). Such a mental setup can lead to implicit biases: For example, a person sincerely acknowledges that she has no prejudices towards Black persons. However, under experimental conditions individuals who have *implicit* prejudices can show discriminatory behavior towards Black persons, regardless of their conscious intentions (for example, see Dovidio & Gaertner 2000; Rooth, 2010).

While implicit biases are unconscious, implicit attitudes are partly conscious. *A characteristic case of implicit bias is an unintended, automatic, and unconscious*

*effect of a psychological mechanism.*³ Implicit attitudes are mental states which are characteristically associated with implicit biases, although psychological entities besides attitudes can, in principle, cause implicit biases as well. Implicit bias “is a term of art referring to evaluations of social groups that are largely outside of conscious awareness or control” (Brownstein & Saul, 2016, p. 1).

When psychologists started to intensify research on implicit attitudes, it was widely considered that implicit attitudes were unconscious mental states. Greenwald & Banaji (1995) defined implicit attitudes as “introspectively unidentified (or inaccurately unidentified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects” (p. 8). However, with more empirical studies on implicit attitudes, researchers question the attribute ‘unconscious’. Especially, Gawronski et al., (2006) called into question whether implicit attitudes are really unconscious. It turned out that many participants are able to self-report the implicit attitude’s content which speaks in favor of the participant’s awareness of implicit attitudes (Hahn et al., 2014; Hahn & Gawronski, 2019; Rivers & Hahn, 2019).⁴ However, it is widely doubted that individuals have awareness of the implicit attitude’s effect (impact awareness; for example, see Gawronski et al., 2008). Accordingly, subjects are unaware of the implicit attitudes’ effects on their thoughts and behavior.

3 Motivation and control

In this section, I argue that the internal motivation to control prejudiced reactions is relevant for decreasing implicit biases’ influences. As I will show in the next section, this is puzzling if control would presuppose the awareness of implicit biases which are considered unconscious.

Having implicit prejudices or stereotypes does not guarantee that an individual will show discriminatory behavior. Meta-studies show that the overall effects-sizes of implicit attitudes is not overwhelming (Greenwald et al., 2009; Oswald et al., 2015) but small to medium. This does not show that indirect measurements lack predictive powers. In contrast, it shows something that is well-known for explicit attitudes, but was sometimes neglected for implicit attitudes: a single mental state usually fails to predict distinct behaviors under different circumstances (for example, see Wicker 1969). For instance, just because an individual has a positive affective explicit attitude towards meat, it does not follow that the individual eats meat – she can still be a vegetarian. Accordingly, behavior results from various mental states and environmental circumstances. Recently, Brownstein et al., (2020) argued that implicit attitudes’ effects are strongly regulated by different mental states, like motivations

³ Characteristic cases include acts of unintentional discrimination. However, when participants’ implicit attitudes towards a social group are measured with an Implicit Association Test, then participants might notice that it is harder for them to sort a positive word to a certain social group. While this involuntary effect on reaction times counts as an implicit bias, participants are aware of it. However, it is not a characteristic case which is, for instance, intensively discussed in philosophy or psychology.

⁴ Although individuals have content awareness of their implicit attitudes, individuals may fail to classify what they introspectively access as the content of a mental state. Especially, they may doubt that the introspected data indicates an automatic evaluation.

(Fazio & Olson, 2014). However, meta-studies usually do not take such moderating variables into account, because then there would be too few studies available for the meta-study.

Individuals engage in *intrinsically* motivated activities because they find these activities valuable and interesting (Deci & Ryan, 2000). In the absence of social pressure or reinforcing rewards, such as money, internal motivated activities persist and flourish. In fact, receiving rewards for an internally motivated activity lowers the overall performance and motivation to engage in that activity. According to the self-determination theory (Ryan & Deci, 2017), internal motivations build on more basic psychological needs, namely *autonomy* and *competence*. Autonomy is an antagonistic motivational process to reinforcement and satisfies the human need to engage in activities only because they matter to us. The need for competence is the need to develop an ability by overcoming challenging tasks, while feeling responsible for the performance. Overall, agents show the best performances for internally motivated activities.

Agents are motivated for different reasons to control their prejudices. Plant & Devine (1998) developed the *Internal* Motivation to Respond Without Prejudice Scale, measuring the desire to appear unprejudiced to oneself, and the *External* (normative) Motivation to Respond Without Prejudice Scale, which addresses the subject's will to avoid race-related conflicts. According to the MODE-model (Motivation and Opportunity as DEterminants; Fazio 1990), the implicit attitude's predictive powers increase if a subject does not have a motivation to handle the situations effortful or does not have the time to deliberate (for a more recent overview, see Fazio & Olson 2014).

In fact, data (Butz & Plant, 2009; Devine et al., 2002; Plant & Devine, 1998) shows that *intrinsically* motivated individuals have an improved control ability and show less biased behavior, in contrast to extrinsically motivated ones that show more implicitly biased behavior. For example, a study by Plant et al., (2010) shows that persons who are internally motivated to control their prejudiced reactions show less biased behavior during interracial interactions (for similar results, see LaCosse & Plant 2019). In contrast, externally motivated persons appeared more prejudiced during the interracial interaction.

Being intrinsically motivated to control prejudiced behavior means to value egalitarianism and to adjust beliefs, goals and intentions accordingly to this acknowledged value. Extrinsically motivated individuals try to avoid social sanctioning only; they control their behavior because of external social pressure. Having implicit prejudices does not automatically lead to implicitly biased behavior. If individuals are intrinsically motivated to control their prejudices, then the predictive powers of implicit prejudices decrease. On the other hand, subjects who lack any motivation to control their prejudiced reactions are more often influenced by implicit attitudes (for an overview, see Fazio & Olson 2014).

4 Reactive control and implicit biases

In the last section, I referred to studies suggesting that internally motivated agents show less implicitly biased behaviors. Now, I will explain why this is puzzling for reactive conceptions of control. Some philosophers (Levy 2014a, b; Saul, 2013, p. 55) and psychologists (Bargh, 1999, p. 363) presuppose that awareness of influential psychological mechanisms is necessary for control. But what role does consciousness play for control (for example, see Vierkant 2017)?

According to Levy (2014a, b), consciousness enables an agent to be (more) coherent by making information accessible to a variety of the brain's encapsulated subsystems. Some mental states, like stimulus-response patterns stored in memory, work unconsciously. For example, seeing a little spider can automatically make somebody jump away. Such unconsciously processed information are encapsulated as they are not sensitive to different mental states, like the explicit belief that tiny spiders are not dangerous. According to the global workspace model of consciousness (Baars, 1997), conscious information is identical with information that is broadcasted across the brain's encapsulated modules. Thereby, an agent can detect conflicting pieces of information in consciousness. Against this background, propositional attitudes, like beliefs or plans, make it possible for an agent to pursue projects through time because through consciousness agents act in coherent and adaptive ways.

Furthermore, being a unified agent is a condition for being held morally responsible for one's actions. Unified agency, so Levy (2014a), builds on consciousness because *certain* conscious mental states constitute us as moral agents across a timespan. Not every conscious mental state constitutes unified agency but personal-level plans, values, and beliefs do. As mentioned in Sect. 2, agents are conscious of their implicit attitudes' contents. However, implicit attitudes are, in this respect, conscious mental states that have properties which *undermine* unified agency. As long as an agent has incoherent explicit and implicit attitudes, there is a high chance that biased behavior will contradict personal-level plans. Only explicit attitudes, so Levy, have the right properties, namely broad content and norm-sensitivity. Conscious mental states with *broad* content respond to a situation's distinct features. Partly, this depends on whether the mental state in question is capable of conscious syntactical manipulations. Such syntactical rules from formal logic help to prevent contradictions. Thereby, an agent can prevent that a recent plan contradicts an earlier one. Besides such sensitivity to norms of rationality, syntactical rules make agents more sensitive to moral norms. These sensitivities contribute to the agent's reasons-responsiveness and thereby to the agent's control.

Levy (see, Chap. 6 in 2014b; direct control is called "responsibility-level control"), for instance, claims that control presupposes that an agent recognizes whether an action or a mental state's content is consistent with an agent's personal-level attitudes. For Levy, it is necessary for an agent to be in control that the agent is conscious of the action's moral significance, or that the agent is conscious of the mental states' contents (and influences) which give an action its moral significance. When agents are influenced by unconscious psychological mechanisms, agents can only correct their behavior if they are aware of how they are influenced.: "Even if we know, on occasion, that we have implicit attitudes [...] we must be able to detect how these

attitudes influence our information processing, and then hit upon a method of modulating or inhibiting this influence” (Levy, 2016, p. 7). Only if an agent is aware of these aspects, the agent can check for existing conflicts.

Regardless of whether implicit attitudes are dispositional states (Baston & Vosgerau, 2016; Machery, 2016) or mental states with representational content (Mandelbaum, 2016), implicit attitudes have properties that entail dispositions to respond in certain ways. An implicit prejudice might be nothing more than the disposition to keep distance from Arab-Muslim persons. While human agency is unified through consciousness by guiding behavior in the light of plans and projects, implicit prejudices do not reflect the agents’ assessment of what they have reason to do. Agents may unintentionally discriminate against Arab-Muslim persons if the agents unknowingly act on implicit prejudices, in contrast to acknowledged values (Rooth, 2010). This happens if agents cannot recognize that there is a conflict between the triggered automatic behavioral disposition to keep distance and the personal value to be egalitarian.

Levy’s understanding of control is reactive because having control consists of (a) awareness conditions, (b) having the ability to recognize, and (c) resolve conflicts. Such a conception of control puts a lot of pressure on an agent’s self-knowledge because being aware of relevant reasons or stimuli is necessary to recognize conflicts.

If an internal motivation to control prejudiced reactions (Plant & Devine, 1998) leads to less biased behavior, is it possible that internally motivated agents can control their behavior, although they do not know that they have implicit attitudes (see, Sect. 2)? Is it possible that they exercise their control ability, while they have no awareness of the psychological effects that implicit attitudes have? From the perspective of reactive control, the answer is *no*. As Levy (2014b) points out, if an agent is unaware of a conflict between an action and her personal-level values, then the agent cannot exercise control over her behavior. Simultaneously, however, implicit biases are characterized as automatic and unconscious influences on thought and behavior. Reactive control cannot explain successful behavior control of internally motivated subjects, because it presupposes that the agent recognizes a conflict between her mental states or between relevant mental states and an action (see Sect. 1).

I do not intend to argue against Levy’s approach of how consciousness plays a role for agency and control (for a critique, see Vierkant & Hardt 2015). In contrast, by focusing on the proactive control strategy, I want to highlight a different aspect of consciousness for agency and control.

5 Proactive control and empirical data

An agent exercises proactive control before she recognizes a conflict, and this activity contributes to goal-directed behavior (Braver, 2012). Because proactive control does not presuppose conflict detection, it opens up the possibility that an agent’s control ability is independent from the agent’s awareness of certain unconscious influences. In this section, I develop the hypothesis that through proactive control agents can shield themselves from unconscious psychological influences.

An agent engages in proactive control if she actively maintains a goal in consciousness.⁵ This activity leads to top-down effects, like inattentional blindness and modifications of unconscious selective attention. Both effects shed light on how proactive control works. After outlining both of them, I will reflect on consequences for agency.

For proactive control, a subject actively maintains a representation in consciousness. If an agent maintains a goal representation in consciousness, then this conscious content has consequences. According to recent theories of consciousness, like the neuronal workspace theory (Dehaene & Changeux, 2011) or the passive frame theory (Morsella et al., 2016), consciousness plays an important function for adaptive action and is not an epiphenomenon that is without effects. One effect of actively maintaining a goal-representation in consciousness is that some salient stimuli are kept unconscious because they are irrelevant to the current task.

It is hard to define what attention is, but for the current purpose I will stick to the biased competition theory of selective attention (Beck & Kastner, 2009; Duncan, 1996). The theory claims that representations in the visual system are competitive. A representation of an object will be at the expense of other object's representations. The competition of what is represented happens automatically and unconsciously in parallel. Additionally, top-down and bottom-up biasing mechanisms influence the ongoing competition. An agent's goal is a top-down influence which pre-activates the internal 'template' of relevant objects. In this sense, a top-down influence is a biasing signal which helps to resolve the competition. However, bottom-up mechanisms bias the competition as well. Very distinctive perceptual features, like a red square on a gray landscape, can bias selective attention in favor of the related object. Furthermore, the competition is integrated across brain systems. If a visual object gains dominance in representation in one system, it will tend to gain similar dominance in different neuronal systems. Attention, then, results from the competition process which gains dominance across different neuronal systems.

An occurrent goal-representation in consciousness has top-down effects on selective attention. Take the effect of inattentional blindness as an example. This effect occurs if an individual does not perceive a salient stimulus, because she lacks attention. In one study (Simons & Chabris, 1999), participants watch a video that shows persons passing a ball to each other, and the participants have the task to count how often the ball is passed. While the participants focus their attention on the ball while counting, they fail to notice that a person wearing a gorilla suit walks into the middle of the basketball game. The gorilla is in the video for nine full seconds, but many participants are not conscious of it.

Whether a person's attention grasps information depends on the information's salience and its relevance to current goals. Attention, however, does not work like a filter that interrupts brain processing of stimuli at a low level. The filter picture, suggested by Broadbent (1958), is wrong because everyday experience and psychological data speak against it. First, even though a subject does not attend to a conversation behind her back, she will notice it if someone mentions her name. If the brain labels

⁵ Arguably, there are different aspects to proactive control, but I will focus on this facet.

unconscious information as important, it becomes conscious. Furthermore, subliminal priming experiments show that the brain processes unconscious stimuli.

This means that because of proactive control, agents are unaware of some information. However, the brain is a parallel processing system and much of its information processing happens unconsciously. So, although a stimulus is unconscious (inaccessible for an agent), the brain processes the unconscious stimuli. The psychological literature is full of examples that show how the brain processes unconscious stimuli, which influences conscious behavior (Bargh & Morsella, 2008). However, unconscious processing is not a fully passive event which happens to agents. Thereby, the second effect of proactive control, which I want to focus on, becomes important.

Unconscious selective attention determines whether the brain processes information unconsciously. Studies with subliminal priming show that unconscious processing is not a purely passive process. Unconscious processing does not just happen to agents and spread through their brain circuits. In contrast, the agent's strategies and expectations play a role for unconscious processing.

A study (Naccache et al., 2002) shows that subliminal priming is not a purely passive, bottom-up process that operates independent of an agent's attention and goals. It is attention that determines whether the brain processes unconscious information or not. If an agent perceives an unconscious prime at an unexpected moment or an unexpected place, then there are no effects of unconscious processing. For instance, unconsciously processing a stimulus, like a word, predisposes the brain to process the same following stimulus more quickly. This effect, however, does not take place when the *unconscious* stimulus is unattended. If the brain processes unconscious stimuli, the brain has to amplify unconscious stimuli with attention, although the stimuli itself is kept unconscious.

In addition, Kiefer & Brendel (2006) showed that unconscious automatic processes are susceptible to attentional modulation. Two experiments showed that masked semantic priming effects are modulated by attention to the prime. The participant's attention to the prime was modulated by a cueing procedure. In the study, the hypothesis was successfully tested that masked priming (primes which the participants are unaware of) has effects when the participants direct their attention to the stimulus stream in the time window of the masked prime presentation, although the participants were unaware of the primes.

Importantly, conscious intentions affect the orientation of unconscious attention, as a study from Woodman & Luck (2003) shows. Participants had the goal to detect squares while ignoring circles. On a computer screen a square appeared on the right and a circle on the left, but both were masked and thereby not consciously perceivable for the participants. Even though the side of the square were unknown to the participants, their visual attention was attracted to the correct target. This shows that the conscious goal had an influence on what the agent unconsciously attended (for a similar result, see Marti et al., 2012).

Importantly, a study by Wheeler & Fiske (2005) shows malleability of implicit stereotyping, which is modulated by the individual's goals. In contrast to the idea that stereotyping is inevitable and passively happens to agents, the study shows that the agent's goals are relevant for implicit cognition. When participants viewed photographs of Black and White faces and had the goal to *categorize* the faces by age, ste-

reotype effects were monitored with functional magnetic resonance imaging (and a lexical priming task). In contrast, if participants had the goal to decide whether a photographed individual would like a particular vegetable, then there were no implicit bias effects measurable. This study illustrates how a conscious goal leads to top-down effects on implicit cognition and prevents unconscious psychological influences.

In sum, if an agent actively maintains a goal representation in consciousness, then conscious and unconscious attention is modulated by the conscious goal. In terms of conscious attention, goal-relevant stimuli become conscious, while goal-incongruent stimuli are excluded. This can lead to effects like inattentional blindness. Due to this mechanism, irrelevant stimuli do not enter the neuronal workspace and stay unconscious. In addition, a conscious goal influences unconscious attention in a goal-directed way. This latter influence can modulate or disperse automatic and unconscious psychological processes, and this happens without the agent's awareness.

6 Proactive control and agency

In this section, I answer the question of what an agent does when she exercises control. I shall argue that the ability to exercise control consists of reactive *and* proactive control strategies. This will highlight the significance of proactive control strategies, because they, together with reactive control strategies, *constitute* our overall ability to exercise control. Furthermore, I want to explain how *agents engage* in proactive control to shield themselves from unconscious psychological influences. Thereby, consciousness plays an additional role for control to what Levy suggested. I will elaborate in more detail on how proactive control is an aspect of *agency* and something that agents *do*.

As explained in Sect. 4, Levy considers consciousness to be essential for exercising control. First, consciousness enables an agent to be (more) coherent by making information accessible to a variety of the brain's encapsulated subsystems. Second, unified agency, a precondition for being held morally responsible, builds on consciousness because *certain* conscious mental states, such as personal-level plans, values, and beliefs, constitute us as moral agents across a time-span. According to Levy, implicit biases undermine unified agency, because they disrupt plans if the agent does not know how psychological influences affect the mind. As outlined, Levy suggests that consciousness plays a reactive role for control, because through consciousness, agents can check for existing conflicts and resolve them. However, as outlined in Sect. 3, data suggest that some agents have more control over implicit biases than others, which is puzzling if Levy's approach is considered, because agents are not aware of implicit biases (Sect. 2). We are now able explain how an agent can exercise control and circumvent implicit biases. The key is to rely on an additional function of consciousness for control (Sect. 5).

Exercising control depends on the ability to use reactive *and* proactive control strategies at the right time. Agents can be more or less good at choosing the right strategies. This is why I think about the ability to exercise control as a *skill*, in contrast to a limited energy resource of willpower (Baumeister & Tierney, 2012) or a

capacity like working memory. A *skill* approach of control has the advantage that it can account for getting better in exercising control through learning.

Reactive and proactive control are *control strategies* which agents use in everyday situations. If an individual is looking for her car keys, she engages in proactive control by maintaining the goal-representation of finding the keys in consciousness. Maintaining a goal-representation in consciousness leads to selective attention. However, if an agent consciously notices that she lost focus, because she is lost in thought, she can resolve the recognized conflict. If so, she engages in reactive control again.

Arguably, there are more proactive control strategies than maintaining a goal-representation in consciousness. For example, consider overcompensation (Kafka, 1995). Here, agents intentionally bias their own responses in favor of a socially disadvantaged group. Arguably, depending on the agent's plans and values, this strategy leads to unified agency, although there could have been unconscious psychological processes that automatically biased the agent against the disadvantaged social group. In this hypothetical scenario, the agent acted on personal reasons and is not influenced by automatic biases. Additionally, the agent does not need to be aware of psychological influences that threaten her unified agency. Data from Kafka suggests that agents use overcompensation if they think they can contribute to restore social justice. So, choosing to overcompensate does not depend on one's awareness of being prejudiced.

Because some agents are internally motivated to control their prejudiced reactions, they successfully shield their behavior from unconscious psychological influences through proactive control. Maintaining a goal-representation in consciousness leads to attentional shift to goal-related properties, like a perceived person's goal-congruence social role. Thereby, the agent does not process information like a racial affiliation, which could serve as a reason. Contextual effects strongly moderate implicit attitudes (for an overview, see Blair 2002). For example, perceiving a person as a Black individual or perceiving the same person as a team-partner shifts automatic attitudes (Barden et al., 2004). Against this background, shifting one's attention away from stereotypical to more goal-relevant properties potentially has the power to circumvent biasing influences.

In fact, the attentional focus influences which category associations are automatically activated. A study by Mitchell et al., (2003) showed that automatic evaluations of Black female and White males were different, depending on whether the participants were attending to gender or race. When gender was salient, participants evaluated Black females more positively, while participants evaluated White males more negatively. When race was salient, participants evaluated Black females negatively and White males positively. Furthermore, another study (Wittenbrink et al., 2001) highlights that the agent's goal plays a role for the activation of automatic evaluations. In the study, one group judged words as good or bad, while the other judged whether the visual characters were real words or non-words. The study shows that if participants focused on the task to evaluate, they showed higher levels of automatic prejudices compared to those participants who focused meaning. The data supports to the idea that attentional shifts to goal-relevant properties can influence automatic processing (Ito & Tomelleri, 2017; Jones & Fazio, 2010; Wheeler & Fiske, 2005).

One characteristic feature of proactive control becomes essential here: Proactive control does not presuppose conflict awareness between an action and a mental state.

Moreover, there does not need to be a conflict at all. In contrast, when agents value things, they pay more attention to realize value-related goals, and their sensitivity to goal-relevant cues improves. Especially distracting ‘forces’ besides motivational conflict can be a target for the top-down effects of actively maintaining a goal-representation in consciousness. These targets include memory weaknesses, lack of focus, and a lack of effort for archiving a goal. Accordingly, consciousness makes not only conflict detection possible, but the contents of consciousness have direct impacts on capacities, such as attention. By maintaining a goal-representation in consciousness, agents can prevent some cognitive mechanisms from automatically processing information and exclude reasons from running processing mechanisms (see Sect. 5). In this way, an additional function of consciousness contributes to our unified agency.

After outlining this approach, I would like to give complementary explanations by raising and answering two questions. Additionally, to handle a challenge regarding the question of how free we are in using control strategies, I shall explain how proactive control relates to agency.

First, how could it possibly support an agent’s control to *exclude* reasons from processing? Levy follows a reasons-responsiveness account, which suggests that being sufficiently responsive to reasons is a necessary precondition for being in control. For example, if the agent is not sufficiently responsive to reasons, the agent must fail to recognize that certain psychological influences *contradict* personal-level plans.

Arguably, because agents respond to reasons, some of their internal reasons can undermine the agents’ best intentions. Agents’ actual reasons are their internal ones (Levy, 2009): Their reasoning procedures can only process the reasons that are actually processed by the agents’ brains. Sometimes, an agent’s actual reasons are misleading and disruptive.

For instance, implicit attitudes influence the capacity of facial perception (Hugenberg & Bodenhausen, 2003). Agents with implicit prejudices towards people of color have the tendency to misclassify facial expressions of Black persons. It is easy to imagine how a misclassified facial perception can lead to discrimination towards a person of color. This misclassification gives the agent a *reason* for reacting unfavorable, and because the agent is reasons responsive, the agent acts accordingly. Against this background, getting rid of some reasons has the potential to make an agent’s good will effective. Not every represented and processed reason is a correct reason, and proactive control explains how some goal-incongruent reasons, here certain perceptual data, are excluded from processing through attentional selection (not only visual perception but perception in general is influenced by implicit attitudes, see Xiao et al., 2016).

Second, are the majority of situations in which implicit biases have troubling consequences situations in which agents focus consciously on relevant counter content? I think that individuals with nonprejudiced motivations learn to associate egalitarian thoughts with members of stereotyped groups, and that such associations become automatized (Bargh, 1997; Moskowitz et al., 1999). When egalitarian individuals see a stereotyped person, this mechanism sub-activates relevant goal-representations. This, however, does not necessarily mean that the agent consciously focusses on the goal to be unprejudiced the whole time. According to the global workspace model of consciousness, consciousness has a *fringe*. While focal consciousness includes

immediate, detailed experiences, the fringe covers information in working memory which agents can access without being currently able to experience it explicitly in detail (Baars, 1997, p. 44). Regardless whether a representation is in focal or fringe consciousness, the representation is in consciousness and is broadcasted through the neuronal workspace.

Baars (1997) illustrates this point with the following example: Try to memorize the numbers 34, 74, 9, 4, and 21. Afterwards, close your eyes and recall each number. People agree that every number which appears in inner speech or through mental visualization is conscious. However, the fact that we have to focus on one number after another does not mean that all other numbers are not in consciousness. They are in working memory and highly accessible (for the relation of accessibility and consciousness, see Cohen & Dennett 2011); they are in fringe consciousness.⁶ Notice that having the ability to exercise control does not guarantee that we succeed with our performances. There is always the possibility of performance errors, regardless of how complex our control strategies are.

Here is a challenge: If agents have no control about whether they are intrinsically or extrinsically motivated to control their prejudiced reactions, in which sense do they have control over their implicit biases if acquiring the motivation that allows for that control is not up to the agents? Although an internal motivation to be egalitarian is correlated with more efficient means to act egalitarian, an externally motivated agent *can* use the same means. All I suggested is that internally motivated agents are more willing to do so. Having control over beliefs, desires or motivations, call it mental-state-control, is not the type of control agents have (Mele, 1995, p. 225). However, agents do not need such mental-state-control to exercise control over their actions, because the agents' actions are not determined by, say, the agents' desires but on whether agents do or do not act on desires.

The suggested answer to the challenge depends on the relation of agency and the proactive control strategy. More has to be said about this relation. I will show that proactive control is an aspect of agency. This means that proactive control is something agents *do*.

Agents intentionally engage in proactive control strategies. Intentional action results from an executive attitude towards an action plan (Mele, 2017).⁷ The most basic action plan comprises one node which stands in for a single action, like raising one's arm. Usually, action plans are more complex. The action plan to buy groceries, for instance, consists of going to the store, find items, and pay. While agents can have different attitudes towards an action plan, like desiring or being afraid, these attitudes are no executive attitudes. The intention to play tennis leads to actually start playing, and this attitude cannot be substituted with beliefs and desires. One can belief

⁶ Maintaining a goal-representation in consciousness may still sound metaphoric. There is, however, an intuitive understanding of what it means to keep focused on a task over a period of time. For now, I think, this is all that is needed to develop the approach. Nevertheless, I agree that a more fine-grained phenomenological analysis of proactive control strategies is demanded to fully understand how agents successfully exercise proactive control.

⁷ Mele distinguishes between distal and proximal intentions. Proximal intentions are those intentions which are generated temporally close to the intentional action. In contrast, distal intentions function like long-term plans. I think that proactive control is foremost interesting for proximal intentions.

that stop smoking is best, and desire to stop smoking, while having no intention do so. Arguably, there is more to intentional action than having the executive attitude towards an action plan. For instance, having the intention to *A* implies that the acting agent lacks the belief that she will not *A*. I am, however, agnostic on this matter.

Agents execute proactive and reactive control strategies intentionally when the usage of strategies is suitably guided by action plans. According to the Simple View of intentional action, an agent only intentionally does *A* if the agent has the intention to *A*. This approach, however, leads to various problems (Mele & Moser, 1994).⁸ Therefore, Mele and Moser suggest the Complex View: If an agent intentionally *A-s*, the agent intends something to which *A-ing* is suitably related. For example, when an individual has the intention to play tennis, she intentionally engages in subsidiary actions, like intentionally holding the tennis racket. Similarly, if an agent engages in proactive control, the agent does not need the intention to do so. While Braver (2012) claims that agents usually switch between proactive and reactive control to solve a task, this does not presuppose that agents have an intention to choose these strategies. Such an intention would require an action plan itself and an executive attitude towards it.

If an agent engages in proactive control to solve a task, this activity is suitably related to the intention to solve the task in question. Suppose a hiker intentionally takes a hike to a mountain's peak. The action plan comprises driving to the parking lot, putting on different shoes, reach certain intermediate goals. While hiking, however, the agent intentionally balances herself when slipping, although it is not part of the action plan. Intentions have a guidance function because of their intention-embedded plans (Mele, 1992; Mele & Moser 1994). This guidance incorporates basic navigation to successfully reach an action's goal. If this guidance did not exist, it would be hard to imagine how intentional action could be successful. Simultaneously, action cannot count as intentional if the agent does not have the required skills to be successful. Winning in a lottery, for example, is no intentional action, because winning is pure luck (Mele & Moser, 1994).⁹ Analogously, agents intentionally engage in proactive and reactive control because it is suitable related to the action plan. These control strategies are ways to counterbalance thought and action to reach one's goal.

7 Conclusions

Agents who actively maintain a goal in consciousness to solve a task are engaging in proactive control. Proactive control has top-down influences on attention and exclude conscious and unconscious goal-unrelated reasons from cognitive processing. An important difference between proactive and reactive control is the fact that reactive

⁸ One problem, for instance, is a possible regress of intentions. If playing tennis is an intentional action, then every bodily movement while playing tennis can be considered as intentional. If, however, intentional movement presupposes an intention, then every single body movement would require an intentional mental state. This is a very implausible result.

⁹ Because the success rate of engaging in proactive control is in many cases not transparent for agents, the actual success rate could undermine the intentionality of proactive control. In that case, however, agents intentionally *try* to control themselves by proactive control.

control presupposes that agents have awareness of potentially conflicting entities, like awareness of relevant mental states and an adequate awareness of an action. With proactive control, however, agents can engage in a control strategy without detecting a conflict. This opens up the possibility that agents can successfully engage in control strategies, although they are unaware of relevant mental states, like implicit attitudes.

I argued that when agents exercise control over their behavior they use reactive and proactive control strategies. Together with the reactive control strategy, proactive control is an essential way to control behavior and build a unified agency. Thereby, besides making reactive control possible, consciousness contributes in an additional way to agency. The proactive control strategy is an agential activity because agents intentionally use it.

Acknowledgements I am especially grateful to Michael McKenna from the University of Arizona. Additionally, I would like to thank Caleb Dewey, Benjamin Keoseyan, and Robert H. Wallace. Also, I thank Albert Newen from the Ruhr University Bochum, as well as Sabrina Coninx, Roy Dings, and Julia Wolf.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Research Foundation [grant number BA 6721/1-1].

Conflict of Interest René Baston declares no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baars, B. J. (1997). *In the Theater of Consciousness: The Workspace of the Mind* by Bernard J. Baars. Oxford University Press
- Banaji, M. R. (2001). Implicit Attitudes Can Be Measured. In H. L. Roediger III, J. S. Nairne, & I. Neath (Eds.), *The Nature of Remembering: Essays in Honor of Robert G. Crowder* (pp. 117–150). American Psychological Association
- Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, 87(1), 5–22. <https://doi.org/10.1037/0022-3514.87.1.5>
- Bargh, J. A. (1997). The automaticity of everyday life. *The automaticity of everyday life: Advances in social cognition* (10 vol., pp. 1–61). Erlbaum
- Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). Guilford Press
- Bargh, J. A. (2017). *Before You Know It: The Unconscious Reasons We Do What We Do*. Atria Books
- Bargh, J. A., & Morsella, E. (2008). The Unconscious Mind. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 3(1), 73–79
- Baston, R., & Vosgerau, G. (2016). Implicit attitudes and implicit prejudices. *Philosophical Psychology*, 29(6), 889–903

- Baumeister, R. F., & Tierney, J. (2012). *Willpower: Rediscovering the Greatest Human Strength*
- Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, 49(10), 1154–1165. <https://doi.org/10.1016/j.visres.2008.07.012>
- Blair, I. V. (2002). The Malleability of Automatic Stereotypes and Prejudice. *Personality and Social Psychology Review*, 6(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8
- Braver, T. S. (2012). The variable nature of cognitive control: A dual-mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113
- Broadbent, D. (1958). *Perception and Communication*. Pergamon Press
- Brownstein, M., Madva, A., & Gawronski, B. (2020). Understanding Implicit Bias: Putting the Criticism into Perspective. *Pacific Philosophical Quarterly*, 101(2), 276–307
- Brownstein, M., & Saul, J. (2016). Introduction. In M. Brownstein, & J. Saul (Eds.), *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology* (pp. 1–19). Oxford University Press
- Butz, D. A., & Plant, E. A. (2009). Prejudice control and interracial relations: The role of motivation to respond without prejudice. *Journal of Personality*, 77(5), 1311–1341
- Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, 15(8), 358–364
- Custers, R., & Aarts, H. (2010). The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness. *Science*, 329(5987), 47–50
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8(4), 377–383
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268
- Dehaene, S., & Changeux, J. P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, 70(2), 200–227
- Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, 82(5), 835–848
- Dovidio, J. F., & Gaertner, S. L. (2000). Aversive Racism and Selection Decisions: 1989 and 1999. *Psychological Science*, 11(4), 315–319
- Duncan, J. (1996). Cooperating brain systems in selective perception and action. In I. Toshio, & J. L. McClelland (Eds.), *Attention and performance 16: Information integration in perception and communication* (pp. 549–578). The MIT Press
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (23 vol., pp. 75–109). Academic Press
- Fazio, R. H., & Olson, M. A. (2014). The MODE Model: Attitude-Behavior Processes as a Function of Motivation and Opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind*. Guilford Press
- Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When “just say no” is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Psychology*, 44, 370–377
- Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are “implicit” attitudes unconscious? *Consciousness and Cognition*, 15(3), 485–499
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27
- Greenwald, A. G., McGhee, D., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41
- Haddock, G., & Maio, G. R. (2015). Attitudes. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.), *An Introduction to Social Psychology* (6th edition, pp. 171–200). Bps Textbooks in Psychology
- Hahn, A., & Gawronski, B. (2019). Facing one’s implicit biases: From awareness to acknowledgment. *Journal of Personality and Social Psychology*, 116(5), 769–794
- Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology General*, 143(3), 1369–1392
- Holroyd, J. (2012). Responsibility for Implicit Bias. *Journal of Social Philosophy*, 43(3), 274–306

- Holroyd, J., Scaife, R., & Stafford, T. (2017). What is implicit bias? *Philosophy Compass*, 12(10), e12437
- Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, 14(6), 640–643. <https://doi.org/10.1046/j.0956-7976.2003.psci.1478.x>
- Ito, T. A., & Tomelleri, S. (2017). Seeing is not stereotyping: The functional independence of categorization and stereotype activation. *Social Cognitive and Affective Neuroscience*, 12(5), 758–764. <https://doi.org/10.1093/scan/nsx009>
- Jones, C. R., & Fazio, R. H. (2010). Person Categorization and Automatic Racial Stereotyping Effects on Weapon Identification. *Personality and Social Psychology Bulletin*, 36(8), 1073–1085. <https://doi.org/10.1177/0146167210375817>
- Kafka, P. (1995). *Low prejudiced people, their ideals, and outgroup overcompensation*. http://digitool.library.mcgill.ca/R/?func=dbin-jump-full&object_id=40153&local_base=GEN01-MCG02
- Kiefer, M., & Brendel, D. (2006). Attentional modulation of unconscious “automatic” processes: Evidence from event-related potentials in a masked priming paradigm. *Journal of Cognitive Neuroscience*, 18(2), 184–198
- LaCosse, J., & Plant, E. A. (2019). Internal motivation to respond without prejudice fosters respectful responses in interracial interactions. *Journal of Personality and Social Psychology*, 119(5), 1037–1056
- Levy, N. (2009). Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick. *Ethics*, 119(4), 729–741. <https://doi.org/10.1086/605018>
- Levy, N. (2014a). Consciousness, Implicit Attitudes and Moral Responsibility. *Noûs*, 48(1), 21–40
- Levy, N. (2014b). *Consciousness and Moral Responsibility*. Oxford Univ Pr
- Levy, N. (2016). Implicit Bias and Moral Responsibility: Probing the Data. *Philosophy and Phenomenological Research*, 94(1), 3–26
- Machery, E. (2016). De-Freuding Implicit Attitudes. In M. Brownstein, & S. Jennifer (Eds.), *Implicit Bias and Philosophy* (1 vol., pp. 104–129). University Press
- Mandelbaum, E. (2016). Attitude, Inference, Association: On the Propositional Structure of Implicit Bias. *Noûs*, 50(3), 629–658
- Marti, S., Sigman, M., & Dehaene, S. (2012). A shared cortical bottleneck underlying Attentional Blink and Psychological Refractory Period. *Neuroimage*, 59(3), 2883–2898
- Mele, A. R. (1995). *Autonomous Agents: From Self-Control to Autonomy*. New). OUP
- Mele, A. R. (2017). *Aspects of Agency: Decisions, Abilities, Explanations, and Free Will*. Oxford University Press
- Mele, A. R., & Moser, P. K. (1994). Intentional Action. *Noûs*, 28(1), 39–68
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, 132(3), 455–469. <https://doi.org/10.1037/0096-3445.132.3.455>
- Morsella, E., Godwin, C. A., Jantz, T. K., Krieger, S. C., & Gazzaley, A. (2016). Homing in on consciousness in the nervous system: An action-based synthesis. *Behavioral and Brain Sciences*, 39
- 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(1), 167–184. <https://doi.org/10.1037/0022-3514.77.1.167>
- Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious Masked Priming Depends on Temporal Attention. *Psychological Science*, 13(5), 416–424
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, 108(4), 562–571
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811–832
- Plant, E. A., Devine, P. G., & Peruche, M. B. (2010). Routes to Positive Interracial Interactions: Approaching Egalitarianism or Avoiding Prejudice. *Personality and Social Psychology Bulletin*, 36(9), 1135–1147
- Rivers, A. M., & Hahn, A. (2019). What Cognitive Mechanisms Do People Reflect on When They Predict IAT Scores? *Personality & Social Psychology Bulletin*, 45(6), 878–892
- Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3), 523–534
- Ryan, R. M., PhD, & Deci, E. L. (2017). (PhD. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. Guilford Publications
- Sandis, C. (2012). *The Things We Do and Why We Do Them*. Palgrave-Macmillan

- Saul, J. (2013). Implicit Bias, Stereotype Threat, and Women in Philosophy. In K. Hutchison, & F. Jenkins (Eds.), *Women in Philosophy: What Needs To Change?* (pp. 39–60). Usa: Oxford University Press
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28(9), 1059–1074
- Vierkant, T. (2017). Does science show that we lack free will?. *Philosophy, Science and Religion for Everyone* (pp. 140–148). Routledge
- Vierkant, T., & Hardt, R. (2015). Explicit Reasons, Implicit Stereotypes and the Effortful Control of the Mind. *Ethical Theory and Moral Practice*, 18(2), 251–265. <https://doi.org/10.1007/s10677-015-9573-9>
- Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: Social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, 16(1), 56–63. <https://doi.org/10.1111/j.0956-7976.2005.00780.x>
- Wicker, A. W. (1969). Attitudes versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects. *Journal of Social Issues*, 25(4), 41–78
- Wilson, T. D. (2004). *Strangers to Ourselves: Discovering the Adaptive Unconscious* (1st Edition). Belknap Press
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81(5), 815–827
- Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology Human Perception and Performance*, 29(1), 121–138
- Xiao, Y. J., Coppin, G., & Bavel, J. J. V. (2016). Perceiving the World Through Group-Colored Glasses: A Perceptual Model of Intergroup Relations. *Psychological Inquiry*, 27(4), 255–274

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.