Routledge
Taylor & Francis Group

Philosophical Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cphp20

# In-between implicit and explicit

Anna Strasser

**To cite this article:** Anna Strasser (2020) In-between implicit and explicit, Philosophical Psychology, 33:7, 946-967, DOI: 10.1080/09515089.2020.1778163

**To link to this article:** https://doi.org/10.1080/09515089.2020.1778163

PHILOSOPHICAL PSYCHOLOGY
2020, VOL. 33, NO. 7, 946–967
https://doi.org/10.1080/09515089.2020.1778163

Routledge
Taylor & Francis Group

ARTICLE



# In-between implicit and explicit

Anna Strasser

Berlin School of Mind and Brain, Humboldt-Universitat zu Berlin, Berlin, Germany

**ABSTRACT**

Research in social cognition aims to illuminate how agents can understand, communicate, and interact with other agents. When defining socio-cognitive abilities, standard cognitivist approaches tend to require demanding representational information processing. Thereby, they describe rather ideal cases. However, interdisciplinary research indicates multiple forms of how socio-cognitive abilities can be realized. Recent minimal approaches offer notions accommodating different kinds of cognitive processing. Nevertheless, the introduction of minimal cases of cognition raises new questions of how to account for commonalities and differences with respect to the standard concepts. It seems to be a widespread strategy to adapt ideas of a two-system approach in order to distinguish less demanding instances from more demanding cases. This paper critically explores such an interpretation of a two-system approach and argues that a dichotomous understanding fails to capture the actual diversity of cognitive processes.

**ARTICLE HISTORY**
Received 26 October 2018
Accepted 19 November 2019

**KEYWORDS**
Social cognition; two-system approach; diversity of cognitive processes

## 1. Introduction

Aiming toward a clarification of notions and developing as clear-cut definitions as possible are part of the major objectives of philosophy. However, in aiming to capture a broad spectrum of a phenomenon with all its different instances, it seems unavoidable that definitions become less sharp, whereas striving for clear-cut definitions often results in rather restrictive notions that cover only one single demanding instance of a phenomenon. Consequently, other less demanding instances of the same phenomenon get neglected. This might as well be due to the tradition to rather over-intellectualize cognition in general. That means it can be observed that there is a tendency to opt for more sophisticated conditions than necessary. Turning to standard definitions of socio-cognitive abilities such as mind-reading (Fodor, 1987; Gopnik, 2003), individual agency (Davidson, 1980), or joint actions (Bratman, 2014), one can summarize that standard notions describe rather ideal cases. Such definitions require demanding cognitive

**CONTACT** Anna Strasser annakatharinastrasser@gmail.com Berlin School of Mind and Brain, Humboldt-Universitat zu Berlin, Berlin, Germany

resources which are often not necessarily in play when a particular socio-cognitive ability is performed. It seems to be a general weakness of such philosophical conceptions that a broader spectrum of cognitive processes is not considered. However, interdisciplinary research results pointing, for example, to abilities of children (Brownell, 2011), nonhuman animals (Warneken et al., 2006), as well as performances of human adults provide a strong motivation to account for various instances in a broader spectrum, rather than just for one single full-fledged and supposedly paradigmatic instance of a socio-cognitive ability.

To remedy this shortcoming, current so-called minimal approaches develop notions that capture alternative realizations of socio-cognitive abilities (Butterfill & Apperly, 2013; Michael et al., 2016; Pacherie, 2013; Strasser, 2018; Vesper et al., 2010). For example, the notion of minimal mindreading (Butterfill & Apperly, 2013) captures an automatic form of mindreading found in human adults, infants, nonhuman animals, and possibly in artificial agents. In the same vein, by proposing a notion of "shared intention lite," Pacherie (2013) is able to call the necessity of Bratman's demanding conditions concerning shared intentions into question. Unlike standard conceptions, which require intentionality, consciousness, and controlled processes, these new notions accommodate unintentional, unconscious, automatic ways of processing. Introducing minimal cases of cognition helps to overcome the restrictions of standard notions and suggests an extension of the realm of cognition.

Even though minimal notions offer an important extension and open up the possibility of capturing multiple realizations of a phenomenon, they are not meant to replace standard notions. Therefore, introducing minimal notions raises questions about how minimal and standard notions relate to each other. A widespread strategy to distinguish demanding from less demanding ways of processing refers to dual-process theories or the two-system approach (Evans & Stanovich, 2013; Kahneman, 2011). As a paradigmatic example, this paper will investigate how Butterfill and Apperly (2013) use the idea of a two-system approach to characterize the underlying operations of minimal and full-fledged mindreading. In a nutshell, they claim that processes realizing minimal mindreading meet the conditions of a typical System 1 process: they are automatic – minimal mindreaders are neither aware of them, nor do they have control over them. By way of contrast, it is assumed that full-fledged mindreading is best characterized by properties of System 2 processes such as being nonautomatic, accessible to consciousness, and controlled.

At first glance, explaining multiple realizations of one and the same socio-cognitive ability by reference to distinct systems appears to be an attractive strategy. System 1 seems best equipped to capture cognitively less demanding and less effortful processes. Apart from being automatic, unconscious,

and uncontrollable, System 1 does not require demanding cognitive resources such as mastery of language, higher-order representations, and others. In contrast, System 2 processes rely on nonautomatic, controllable processes and require, for example, mastery of language, and meta-representations. Thus, they are cognitively demanding.

This paper aims to critically question whether the way in which a two-system approach is used to distinguish minimal from full-fledged realizations results in an appropriate distinction between their underlying operations. To this end, I argue for the general claim that categorizing cognitive processes into two distinct systems leads to a questionable implicit–explicit dualism. With reference to empirical evidence, I show that a dichotomous interpretation of the two-system approach fails to capture the diversity of cognitive processes. Consequently, I suggest taking the idea of a two-system approach simply to mark extreme points in a wide spectrum of various property distributions. By setting up a continuum, one can better facilitate in capturing a variety of in-between cases.

In order to explore the consequences of ascribing properties to cognitive processes according to a dichotomous interpretation of a two-system approach, I first present potential properties of cognitive processes, and then specify the properties which are used to characterize a two-system approach (Section 2). Analyzing a dichotomous interpretation of a two-system approach, I argue that, contrary to a common usage of a two-system approach, properties of cognitive processes vary in degrees, not in supposedly basic types. Furthermore, I reject the claim that properties characterizing one of the two systems necessarily co-occur. Based on empirical evidence, I demonstrate that there are other potential combinations of properties that combine properties of both systems (Section 3). Finally, focusing on the question as to whether processes coming from distinct systems might influence each other, I discuss several examples that provide an objection to the claim of informational encapsulation of System 1 processes and show that processes from both systems can mutually influence each other (Section 4).

## 2. Properties of cognitive processes

Without a doubt, there are multiple realizations of socio-cognitive abilities. Children's abilities present a paradigmatic example since they are obviously capable of several socio-cognitive abilities long before they fulfill the demanding conditions of ideal cases described by standard notions (Brownell, 2011). The development of minimal notions introduces a conceptual framework of how to describe less demanding realizations of socio-cognitive abilities (Butterfill & Apperly, 2013; Michael et al., 2016; Pacherie, 2013; Strasser, 2018; Vesper et al., 2010). However, pointing to

different kinds of cognitive processes which account for demanding and less demanding realizations of a socio-cognitive ability, we will also have to specify the properties of the underlying operations by which we can account for differences and commonalities.

Up to now, there are various disagreements when it comes to clarifications about specific features of cognitive processes. For example, with respect to boundaries of cognition, irreconcilable positions face off against one another. Some claim that cognitive processes are necessarily internal and brain-bound, whereas others argue that cognition should be understood as extended into the body and environment. Furthermore, there are debates about the question of whether associative conditioning and other seemingly lower-level processes may count as cognitive (cf. Buckner & Fridland, 2017). On the one side, we have representational theories of the mind, such as the Computational Theory of Mind (CTM), which explain cognition by reference to mental representations and computations (Thagard, 2019). On the other side, there are positions, such as enactivist approaches, that argue for skepticism concerning mental representation and computation (Brooks, 1991; Hutto et al., 2011). Nevertheless, there seems to be a consensus that cognitive processes are based on some sort of information processing that enables systems to anticipate, decide, act, and much more.

For the aim of this paper, I refer to cognitive processes as input–output relations. Even if this might be too much of a simplification, I think this is still a useful framework. To cover the nature and diversity of cognitive processes in depth, a more complex model would be needed. Such a model should not only integrate embodied, embedded, enacted, and extended factors, but also capture the reciprocal dynamics of cognitive processes. It would also be desirable if such a model were able to cover both representative and non-representative explanations of intelligent behavior. However, for the aims of this paper, namely, to present a critique of dichotomous distinctions, this simplified input–output model may be sufficient. Nevertheless, I hope that even positions that profoundly contradict an input–output framework can benefit from my considerations by abstracting from the description used here.

Under the assumption that cognitive processes can be understood as input–output relations which expand over time, I distinguish three aspects of such processes, describing distinct phases of the manner(s) in which inputs are transferred to outputs:

(1) Properties concerning possible input parameters
(2) Properties characterizing intermediate operations
(3) Specifications ascribing properties to possible outputs

One cognitive process can possess different properties concerning each of these three aspects. For instance, verbal inputs may qualify as input while

intermediate operations proceed with nonconceptual representations, and the properties describing the output may be restricted to looking behavior. Properties characterizing input parameters specify what kind of stimuli can start a process, and thereby, can also give first insights about what type of information is accessible for a process. Such properties may serve as an indicator, for example, for domain-specific processes. Leaving characterizations concerning the input and the output aside, I assume that multiple realizations of a socio-cognitive ability are likely to differ, above all, in their intermediate operations. Given that the characteristics of input and output are similar, multiple realizations open up the possibility that the intermediate processes still differ with respect to properties that contribute to factors concerning controllability, structure, and accessibility. Accordingly, the potential properties of intermediate operations will stand in the center of this investigation.

To account for controllability, one has to investigate the relations between the intermediate operation and the rest of the cognitive system. Having control over a process includes capacities to intervene. One important condition for being able to initiate changes involves the question of accessibility of information. Some operations are fully automatic; once the stimulus is perceived, the subsequent operations are started and cannot be inhibited. Other operations can be subject to voluntary control and can be stopped or manipulated, even if the operation has already commenced; and, as I argue in this paper, there are also operations that are more-or-less automatic.

Evaluating the structure of cognitive processes, we can distinguish complex from rather simple structures. Some structures require a lot of resources, such as intense memory capacities or sophisticated representational formats, while others are rather self-sufficient and potentially more efficient. The nature of the structure influences the performance of intermediate operations, which can be described by relational properties such as speed, flexibility, or robustness; and as processes expand over time, it is also possible that properties of the intermediate operations may change over time.

Especially central to this paper are properties concerning questions of accessibility. In analyzing accessibility, we have to consider two things.

On the one hand, information used by an intermediate operation may be accessible to other processes, thus making such processes cognitively penetrable. To this end, I distinguish cases where we are aware of using information – and are even able to report this – from cases in which information is used but we are not aware that this information being used. In the latter case, information is only accessible in an implicit manner. This means other cognitive processes may make use of some information, but the person is not aware of this. In such a case – when "central

monitoring" has no access – we speak of limited central accessibility. In view of an interesting distinction made by phenomenologists (Frank, 2015; Gallagher & Zahavi, 2007, 2008), namely the difference between pre-reflective and reflective forms of consciousness, one can also describe cases in which we are somehow aware of using information without being aware of being aware.

On the other hand, intermediated operations realizing a socio-cognitive ability may have access to information stored elsewhere in the cognitive system. If an operation, in principle, cannot access information which is stored in other parts of the cognitive system, it is informationally encapsulated. Even though it seems possible that these two directions of accessibility can come apart, they are often taken as co-occurrent. Robbins (2017) framed these two directions, namely, informational encapsulation and limited central accessibility, as the two sides of the same coin. If intermediate operations cannot access information which is stored elsewhere, and other processes cannot access information used by these intermediate operations, one can conclude that such operations can neither be inhibited nor modulated. These operations seem to be immune to any influences; or, in other words, we are at the mercy of such processes.

Gaining epistemic access to the concrete properties of intermediate operations, realizing a specific socio-cognitive ability, presents a particular challenge to empirical research, since such properties tend to hide in the black box between input and output. As a matter of fact, not all intermediate operations are accompanied by consciousness. Of course, one should assume several levels of consciousness, since just distinguishing conscious versus unconscious falls back in a rather dichotomous distinction which is criticized in this paper. Therefore, exclusively relying on verbal reports is not sufficient. Moreover, without resulting in easily measurable differences, distinct intermediate operations can lead to the same output. So far, we are often unable to distinguish the various formats of representations which are used by an intermediate operation. Due to methodological difficulties of directly measuring properties of intermediate operations, many experimental settings face the so-called logical problem; namely, that conflicting interpretations can provide explanations of equal value for the same outcome. Consequently, the ascriptions of properties concerning intermediate operations often suffer from uncertainty because multiple realizations can potentially produce the same outcome. This can be seen most prominently in the debate about animal mindreading, in which behavioral hypotheses compete with mindreading hypotheses (Lurz, 2011). Up until now, neither side of the debate has been successful in proving the other wrong; but that shouldn't keep us from investigating the potential properties of intermediate operations.

### 2.1. How a two-system approach is used to categorize cognitive processes

The question I want to investigate in this paper is whether referring to a two-system approach (Kahneman, 2011) provides an appropriate framework for characterizing properties of underlying intermediate operations of minimal and full-fledged socio-cognitive abilities.

Accordingly, I take minimal mindreading, introduced by Butterfill and Apperly (2013), as a paradigmatic example of how this can be done. In a nutshell, a two-system approach distinguishes two systems of reasoning, each of which is characterized by a bundle of properties. Operations assumed to occur in System 1 are thought of as implicit, nonverbal, rapid, automatic, unconscious, and not controlled – they comprise unconscious reasoning, whereas in System 2, we find conscious reasoning. Such operations are described as explicit, linked to language, flexible, slower, rule-based, effortful, controlled, and capable of being inhibited.

Roughly speaking, Butterfill and Apperly claim that full-fledged mindreading is based on rich cognitive resources and cognitively more demanding operations that can be controlled and inhibited. In particular, the handling of complex mental states requires memory capacities, the ability to build up meta-representations, mastery of language, and other executive functions. In order to count as a full-fledged mindreader, you have to be able to recognize that others can have beliefs about the world which are divergent from the ones you have yourself. In other words, one needs to understand that information about one and the same object can differ, depending on the perspective, time of observation, or functional preference of an agent. This resonates quite well with common descriptions of System 2 processes. On the other side, Butterfill and Apperly suggest that minimal mindreading is best described by properties of System 1. Their minimal approach specifies the very minimal presuppositions of mindreading. For example, instead of requiring representations of a wide range of complex mental states, they claim that less demanding mental states, namely encounterings and registrations, are sufficient. According to them, the intermediate operations realizing minimal mindreading are automatic, fast, robust, and cannot be inhibited, thereby qualifying as typical System 1 processes.

In the following, I critically discuss how the two-system approach is used to characterize two types of cognitive processes. I argue that due to a rather dichotomous interpretation, only extreme cases are captured. Even though characterizations of System 1 are often summarized as being implicit, whereas System 2 processes are thought to be explicit, we do not have a universal definition of what it means to characterize operations as implicit or as explicit (Evans, 2008). All we have is a bundle of properties characterizing the processes of each assumed

Table 1.

| ASPECTS OF PROCESSES | | | SYSTEM-ONE UNCONSCIOUS REASONING | SYSTEM-TWO CONSCIOUS REASONING |
|---|---|---|---|---|
| INPUT | | | domain-specificity | diverse input parameters |
| INTER- MEDIATE OPERATIONS | CONTROL | | no voluntary control, unintentional | voluntary control, intentional |
| | | | automatic | non-automatic |
| | SPEED | | fast | slow |
| | ACCESSIBILITY | CENTRAL | not available to consciousness | available to consciousness |
| | | OTHER INFORMATION | not accessible | accessible |
| | | FOR OTHER PROCESSES | information is not accessible | information is accessible |
| | STRUCTURE | | simple computational operations | effortful, cognitively demanding |
| | | | hardwired, fixed neural architecture, robust | adaptive, flexible |
| | DEVELOPMENTAL FACTORS | | innate or developed early | developed later |
| OUTPUT | | | looking and other sense specific behavior | verbal reports |

system. The list in Table 1 is not exhaustive, but this selection of properties can serve as an example of how properties are grouped in a rather dichotomous framework.

Table 1 lists properties which are frequently used to distinguish System 1 from System 2 processes. At first sight, this resonates with the proposed description of minimal and full-fledged mindreading by Butterfill and Apperly (2013). However, reflecting on the way in which processes are distinguished, it becomes obvious that this view inherits several critical shortcomings. First, these properties are presented in a dichotomous way, which suggests an either–or ascription, and which excludes a variation by degrees. Second, it is important to mention that it is tacitly assumed that these properties necessarily co-occur with each other. This means that it is assumed that the properties characterizing one system stand in an all-or-none relation.

Taking the above table as a template, the left column should provide the defining characteristics of a prototypical System 1 process. For example, the input is characterized by a domain-specific input channel. The intermediate operations are neither subject to direct voluntary control nor available to consciousness. They produce their outputs relatively quickly without being able to make use of other information in the cognitive system, and, in addition, other processes cannot access information about these operations. In sum, they are thought to be informationally encapsulated and

inaccessible to central monitoring. From a developmental perspective, it is likely that such processes are innate and hardwired, or developed very early. The outcomes happen on a bodily level and do not require a mastery of language. As previously noted, this resonates quite well with the characterization Butterfill and Apperly give for minimal mindreading.

I do not deny that there are processes which qualify as prototypical System 1 processes. For example, a reflex (even though not cognitive) or a visual illusion (like the Müller-Lyer illusion) might fit this description. Sometimes we are at the mercy of a hard-wired reaction to a stimulus and have no chance to inhibit or modulate subsequent processing. Following the idea of the modularity of mind, introduced by Fodor (1983), one might claim that much of our input systems are informationally encapsulated. However, I claim that this is highly questionable. If one were to follow Fodor, one would have to deny that vision can be cognitively penetrable all the way down, and therefore, one would, for example, have to claim that low-level visual processing is in principle informational encapsulated. However, empirical research has shown that vision is cognitively penetrable. Top-down processes can have a significant influence on visual processes which are influenced by an agent's motivational states (e.g., desirable objects look closer; see Balcetis & Dunning, 2006) or by subjects' beliefs (e.g., racial categorization affecting reports of the perceived skin tone; see Levin & Banaji, 2006). Furthermore, other studies show several cross-modal effects in perception (e.g., the McGurk effect (McGurk & Macdonald, 1976); the double flash illusion (Shams et al., 2000); the rubber hand illusion (Botvinick & Cohen, 1998). Such findings speak against informational encapsulation.

I suggest taking the above characterizations of System 1 properties (the left column) just as a description of an extreme case of a rather broad spectrum of possible property distributions. In contrast to Butterfill and Apperly, I claim that minimal mindreading is not such an extreme case. If we want to accommodate the varieties of cognitive processes, we should be cautious with dichotomous characterizations, and we will see that extreme cases of System 1 or System 2 are less frequent than expected.

## 3. Capturing the varieties of cognitive processes

In analyzing a dichotomous interpretation of a two-system approach, I first question the either–or ascription of properties by referring to empirical evidence speaking for a continuous development of properties. If properties vary by degrees, we cannot maintain an either–or ascription. For example, we can easily imagine several stages of automaticity – it makes sense to claim that a process can be more or less automatic. Then, I will demonstrate that above-described bundles of properties do not necessarily co-occur.

Regarding mindreading, for example, we could assume that a professional minimal mindreader might gain limited control over the automatic processes. Questioning, for example, the co-occurrence of automaticity and inflexibility, one could also imagine that flexible full-fledged mindreading is interwoven with automatic procedures.

### 3.1. Properties vary by degrees

Properties specifying the two systems tend to be understood as either–or ascriptions. This becomes obvious if one explores, for example, automaticity. A dichotomous interpretation of a two-system approach leads to a strong version of automaticity claiming that processes are either automatic or not automatic. However, taking experimental research into account, one can find contradicting evidence. For instance, Logan (1985) claims that there is evidence that automaticity is learned, and infers from this that attributions of automaticity are relative judgments. Therefore, automaticity should be viewed as a continuum. Shiffrin (1988) also argues that it is more likely that automatism improves gradually with consistent practice. In the same line, MacLeod and Dunbar (1988) argue for a continuum of automaticity. They showed, in a series of experiments comparing performances in a modification of a Stroop test (Stroop, 1935), that differences concerning potential interferences can be explained as a direct consequence of training. The test persons were not at the mercy of a hardwired automatic reaction to the stimuli. Last but not least, even Kahneman, who originally introduced the two-system approach, argues for a weak version of automaticity capturing gradients of automaticity (Kahneman & Chajczyk, 1983). This is the first hint that Kahneman himself does not stand for a dichotomous interpretation of the two-system approach.

Questioning the sharp separation of cognitive processes into two distinct systems and arguing for a gradual development of properties, Wojnowicz et al. (2009) put forward the thesis that explicitly expressed attitudes can be understood as the result of complex, nonlinear, time-dependent processes in which several less explicit attitudes compete with each other. According to this view, being implicit is not a matter of being generated in a specific system, but rather a question of losing the competition quite early. Introducing this temporal dimension, it seems likely to assume a 'more or less' (i.e., a continuum) instead of a 'yes or no.' Wojnowicz et al. (2009) could show in a study investigating people's hand-movement trajectories for explicitly evaluating "Black people" and "White people" that racial biases were visible in the dynamics of the response movement. Instead of finding evidence for clearly distinct explicit decisions, graded motor curvature indicated that explicit attitudes evolve through continuous temporal dynamics during real-time mental processing. Even though much research

in social psychology takes implicit attitudes as operating outside of cognitive control and, perhaps, introspective awareness (Bargh & Chartrand, 1999), there are examples which indicate that implicit attitudes can be cognitively penetrable. Often, we are able to overcome implicit biases and thereby decrease the influence of implicit attitudes.

Nevertheless, there still seems to be certain kinds of implicit processes which never have the chance to reach a conscious level. Taking this into account, I suggest marking those cases as extreme cases which are not in the center focus of the considerations of this paper. Having said this, I claim that a lot of empirical research speaks for the idea that we are not switching from pure implicit processes to pure explicit processes, it is much more likely that there are continuously competing processes from which, in the end, the behavioral choice emerges. Roughly speaking, one may state that the longer a process survives the assumed competition, the less implicit this process becomes.

### 3.2. Properties do not necessarily co-occur

A further critical consequence of a dichotomous interpretation of a two-system approach consists in the claim that all properties assigned to one system necessarily co-occur. According to this view, automaticity is often seen as co-occurring with four other properties, namely, unconsciousness, unintentionality, efficiency, and uncontrollability (cf. Bargh, 1994). However, empirical research questions whether all four properties necessarily co-occur. Reviewing empirical findings, we find no evidence for an all-or-none relation: to the contrary, there are several counterexamples found in social psychology. Cognitive processes display a combination of properties, making them at the same time automatic in one sense and nonautomatic in another sense. Processes can be conscious but uncontrollable, unintentional but still controllable, or efficient and intentional (Gawronski & Bodenhausen, 2011). The continuum view claims that properties of processes can gradually change. Now, we can add that they can also be disassociated.

For instance, it is often assumed that implicit processes are, in principle, unconscious. However, experimental research about attitudes indicates that one can have experiential access to so-called implicit processes. Consequently, being inaccessible is not a necessary condition of implicit processes. Several studies show that correlations between implicit and explicit evaluations increase when participants are instructed to focus on their feelings for the object of their attitude (Gawronski et al., ; Nier, 2005; Ranganath et al., 2008; Smith & Nosek, 2011). If implicit processes would be, in principle, unconscious, such introspection instructions should leave correlations between explicit and implicit evaluations unaffected (for

a detailed review, see Gawronski & Bodenhausen, 2006). Therefore, one can claim that not all implicit processes are necessarily unconscious or inaccessible.

Consider, for example, your heartbeat, which typically operates outside of your control or awareness. However, these properties are not invariable. It is possible to bring your heartbeat to a level of consciousness, but this does not mean that you automatically succeed in controlling your heartbeat. Even though some people might even succeed in controlling their heartbeat by controlling their breathing or applying certain meditation techniques, there is a stage where they may be conscious of it but not yet able to control it (Peng et al., 2004, 1999). In sum, one can claim that mental practice can foster awareness and control of reactions to external stimuli (Balconi et al., 2017). Conversely, conscious and controlled processes can become unconscious without becoming uncontrolled. For example, a skillful piano player is able to automatize processes while retaining the ability to control and monitor them to a great extent.

The above examples show that properties which are characteristic for System 1 can be combined with properties which are used to characterize System 2. Therefore, we can claim that the four horsemen of automaticity (Bargh, 1994) do not necessarily co-occur. Taking this into account, one can argue for a disjunctive conceptualization of automaticity by which a process can be characterized as automatic if it meets at least one of the four criteria (i.e., unconscious, unintentional, efficient, uncontrollable). Such a decompositional classification schema would require more than just two categories, depending on just how independent the variables are. From a theoretical point of view, this already leads to sixteen possibilities. Assuming that properties vary in degree, as explained in the previous section, we must also assume that the individual properties are not dichotomous throughout. This leads to unboundedly many potentially discernible possibilities which clearly point in the direction of a continuum. Future research could start to investigate which of the sixteen possible combinations apply to which socio-cognitive abilities, but should not neglect various in-between cases. Even though it might be more likely that a process that is not controllable is also unconscious, there are other combinations possible which cannot be captured by a dichotomous interpretation of the above table of properties. Since empirical research indicates that the bundles of properties which characterize System 1 vis-à-vis System 2 processes do not necessarily co-occur, I claim that it is wrong to conclude that all automatic processes are necessarily unconscious, unintentional, efficient, and uncontrollable. In refining notions describing so-called System 1 processes, one should consider a disjunctive conceptualization of automaticity and deliver a specification of the criteria, as well as taking into account that each criterion can vary in degree.

In the following, I argue that there are cases of rather implicit processes which are not, in principle, uncontrollable. To this end, I will argue against the claim of informational encapsulation and show how System 1 processes can be influenced by processes which do not have typical System 1 properties.

## 4. Mutual influences

Dichotomous interpretations of a two-system approach tend to deny the possibility of mutual influences between System 1 processes and System 2 processes. Most importantly, it is assumed that information transported by a prototypical System 1 process (an extreme case) is not accessible to any other processes, especially not to processes of central monitoring. The idea of informational encapsulation already contradicts the above-reported empirical evidence which speaks against the necessity that implicit processes are unconscious (Gawronski & Bodenhausen, 2006). However, positions based on the claim that implicit processes are, in principle, not cognitively penetrable infer that such processes cannot be controlled either. In addition, such processes are thought to be informationally encapsulated. That means such processes cannot access and make use of information stored in other parts of the cognitive system. Because System 1 processes, which I label as extreme cases, are taken as informationally encapsulated and not cognitively penetrable, they seem to be radically separated from all other processes in a cognitive system. In other words, they are blind and invisible. Again, we have already seen that there is empirical evidence which indicates that the performance of implicit processes can be based on learning and experience, which, in turn, speaks for a potential influence of other processes. In addition, one can refer to studies also showing that explicit processes can be influenced by rather implicit processes. For example, Pärnamets et al. (2015) demonstrated that even abstract moral cognition is partly constituted by interactions with the immediate environment and is likely supported by gaze-dependent decision processes.

Applying the dichotomous interpretations of a two-system approach to minimal mindreading as a supposedly prototypical System 1 process, it would follow that its operations would be immune to influences from other processes. Indeed, by claiming that minimal mindreading is informationally encapsulated, Butterfill and Apperly exclude any potential influence of System 2 processes on minimal mindreading. In contrast, I claim that there are various kinds of observable influences. I argue that we are not at the mercy of the automatic operations of minimal mindreading or other rather implicit processes in the same way as we are, for example, tricked by visual illusions. By presenting three types of potential interventions and evaluating empirical evidence, I show in the next section that there are

several ways in which so-called System 1 processes can be influenced by other processes.

### 4.1. Input – blocking stimuli by selective attention

Most descriptions of typical implicit processes seem to assume that they are automatically caused by specific stimuli. Moreover, it is assumed that we are at the mercy of such stimuli. However, selective attention is a process of directing our awareness to relevant stimuli while ignoring irrelevant stimuli in the environment. Several studies in psychology clearly indicate that humans display selective attention concerning, for example, acoustic stimuli; this is colloquially termed the cocktail party effect (Broadbent, 1958; Cherry, 1953; Handel, 1989).

Analyzing this ability, one can roughly say that we are able to block some signals whereas other signals are processed with priority, and the priorities are not fixed and rigid. Besides being able to separate relevant input (e.g., speech) from irrelevant input (e.g., background noise), it has been shown that humans are as well able to separate one voice from another. Otherwise, it would be hardly possible to have a conversation in a room in which many people talk at the same time. Of course, it should be noted that people with psychopathological deficits can have great difficulty in exercising selective attention. Regardless of what specific mechanism(s) enable this – we might have specialized filtering systems (Broadbent, 1958), or our focus of attention may block other inputs – what is important, with respect to the claim of this paper, is that our ability for selective attention questions the idea that processing certain signals is, in all cases, a fully automatic procedure which cannot be controlled. Not precluding uncontrolled automatic processes, we may claim that there are more cases than expected which are questioning the idea of completely automatic and stimulus-driven processes. Arguing along the same lines, it can be questioned whether visual processing is completely automatic and stimulus-driven (Töllner et al., 2012).

Taking such research findings into account, it is most likely that the process of how visual inputs triggers, for example, mindreading is also not completely automatic. Even though empirical research indicates that the mere presence of other agents seems to be a sufficient stimulus to start minimal mindreading, I claim that selective attention can avoid processing such stimuli. This is not to deny that several studies show that test subjects concerned with perspectival judgments tend to slow down and make more mistakes when other agents are present than when other agents are absent (Kovács et al., 2010; Qureshi et al., 2010; Samson et al., 2010). This difference in performance is explained by so-called altercentric interferences caused by minimal mindreading. The mere fact that the participants simultaneously represent two different perspectives of the same situation –

namely, their own perspective as well as the perspective of the other – can lead to interferences. To solve so-called false-belief tasks, it is essential that mindreaders process distinct representations that contain contrasting information; but, in solving other perspectival tasks, having contrasting representations is rather disturbing. In other domains, such interference can, for example, be seen in the reaction time of a task; for example, the Stroop effect (Stroop, 1935). Hypothetically, one can imagine an experimental design that could compare a group of participants which have an incentive to apply selective attention with a group which does not. For example, you could instruct one group to count the number of people in addition to the actual task, while the other group should count the number of windows; or, you could add the information to the instructions of one group that the other people have a distracting effect.

Again, I do not deny that some rather implicit processes might happen involuntarily and cause alter-centric inferences. I even admit that there might be certain stimuli which cannot be blocked, and that, in addition, subsequent processes may be based on a fixed relation between a specific stimulus and a subsequent operation that cannot be changed. However, I question whether this is necessarily so. To this end, imagine a situation in which you experience again and again that minimal mindreading processes disturb you and cause you to act in an unfortunate way. As soon as you realize the connection between the presence of the other agents and the fact that you are not that efficient anymore, you will try to concentrate on your primary task. To facilitate this objective, it would be helpful to inhibit minimal mindreading despite the presence of strong stimuli. Due to selective attention, we are able to block verbal inputs in a crowded room. Therefore, I assume that we are as well able to block visual inputs in crowded surroundings. Consequently, we could inhibit minimal mindreading by blocking the visual input. According to the two-system approach, inhibiting operations are characterized as System 2 processes. Therefore, being able to block certain stimuli that inhibit operations assigned to System 1 is an example that System 1 processes are not immune to influences by other processes that are supposedly quite conscious.

### 4.2. Intermediate operations – not immune to education

Distinguishing three aspects of cognitive processes, namely, input parameters, intermediate operations, and characteristics of the output, I focus now on intermediate operations in order to argue for the adaptability of rather implicit processes within new situations.

Imagine a rather fictional situation in which you have been transferred to an entirely new environment with bizarre inhabitants equipped with very unusual systems of perception. Due to these unusual systems of perception, your

minimal mindreading processes continuously fail to recognize what others can or cannot see. According to the presumed property of informational encapsulation, there is no chance of adapting minimal mindreading to a new situation. However, taking cultural adaption as an example, I would argue that presumptively automatic procedures can also be trained to handle new kinds of stimuli. At least in other domains, we have plenty of evidence that even quite established links between an unconditioned stimulus and a conditioned stimulus can become extinct under certain circumstances (Quirk & Mueller, 2008). Therefore, I argue that we should not exclude the possibility that automatic operations of minimal mindreading can be reexamined. A successful adaption to new stimuli can be interpreted as an ability to learn, which would make implicit processes more flexible. Since learning processes resulting in increased flexibility, such as adapting to new stimuli, do not qualify as prototypical System 1 processes, I claim that this is another example of how implicit processes can be influenced by other processes. It may, in some cases, not be very easy to reeducate supposedly automatic operations, but up to now, I have not found compelling reasons that speak against this possibility.

To back up above considerations, I refer to the review by Gawronski and Bodenhausen (2006) which presents several examples which indicate that associative (implicit) and propositional (explicit) processes do not operate in isolation, but mutually interact with each other. Additionally, one can also refer to experimental research investigating people with a lack of explicit memory (Starr & Phillips, 1970). Those experiments show that even though participants cannot rely on explicit processing, they are still able to learn. This indicates that System 1 is able to learn as well.

### 4.3. Output – ignoring the outcome

The last example concerns the outcomes of implicit processes. Even in cases of failing in blocking input signals or adjusting intermediate operations, we are not at the mercy of the results produced. With respect to mindreading, I claim that full-fledged mindreading can overcome intuitions of System 1. For example, consider operations by which full-fledged mindreading attributes fixed stereotypes coming from System 1, and thus System 1 could pollute the outcome of full-fledged mindreading. Even though it is hard to overcome stereotypes, it is possible (Sassenberg & Moskowitz, 2005). Furthermore, research has shown that training can reduce the automatic activation of stereotypes. Kawakami et al. (2000), for example, showed that subjects could reduce stereotype activation after a specific training period in which they learned to activate counterstereotypic information (Gawronski et al., ; Stewart et al., 2010); if we can successfully ignore outcomes of minimal mindreading, this might have an influence on the activation of minimal mindreading. Hypothetically, I assume that continuously ignoring

specific outcomes can decrease the probability of the occurrence of minimal mindreading operations which cause such outcomes.

## 5. Maybe we don't have to follow a dichotomous interpretation

Taking the above examples into account, one should reconsider the claim that implicit processes cannot be influenced by other processes. Due to selective attention, we are able to block stimuli and thereby inhibit certain implicit processes. In specific situations, we are able to educate or adapt particular operations of implicit processes, and we may decrease the likelihood of their occurrence by consistently ignoring their outcomes. Based on the reported empirical evidence, I claim that we should avoid a dichotomous conceptual framework which can only account for extreme cases of System 1 and System 2 processes.

Does this mean that we should reject two-system approaches in general and turn to one-system approaches, as suggested by Leslie et al. (2004), Baillargeon et al. (2010), and Carruthers (2013), in order to characterize the variety of cognitive processes? Maybe so. One advantage of a one-system approach may be that it is easier to argue for a continuity thesis. However, a one-system approach may also have difficulties to grasp the many differences in the diversity of cognitive processes. Despite all the negative consequences of a dichotomous interpretation of a two-system approach, it still remains an attractive framework for characterizing findings and predicting specific failures. For example, the idea that System 2 is ontogenetically – indeed possibly evolutionarily – developed later can explain why some of the more sophisticated realizations of socio-cognitive abilities develop later. If we can interpret the two-systems idea as a continuum instead of a dichotomy, the approach can prove as a useful simplification in order to point at certain areas of an assumed continuum.

Going back to one of the origins of the two-system framework, we can learn from Kahneman's (2011) explicit descriptions of "System 1" and "System 2" as two fictitious characters, not as systems in the standard sense of entities with interacting parts:

> This book has described the workings of the mind as an uneasy interaction between two fictitious characters: the automatic System 1 and the effortful System 2. . . . And of course, you also remember that the two systems do not really exist in the brain or anywhere else. (p. 415)

Furthermore, Kahneman's characterizations indicate that he is neither presupposing informational encapsulation nor arguing for a strict dichotomous grouping of properties. Of course, he characterizes System 1 as automatic, fast, and autonomous; but he also states that System 1 is hardly educable. I think this is an important point: being

hardly educable does not mean that there is no chance to educate operations at all; and, as we have seen above, implicit processes are not immune to education. If operations of System 1 can be educated, then such operations are not informationally encapsulated. There are surely cases in which education is unachievable. For example, being tricked by visual illusions cannot be turned off; no matter how hard we try, we cannot influence the way such basic operations proceed – these operations seem to be uneducable. However, as soon as we know about the effects of the illusion, we do have the chance to change our experience, because our experience can integrate further evaluations. Furthermore, there are cases in which we are able to change automatic reactions. Kahneman (2011) even describes the possibility that "System 2 has some ability to change the way System 1 works" (p. 23). With respect to System 2, operations are described as effortful and slow, but Kahneman explicitly states that they are often guided by System 1, even though it is an achievement of System 2 to overcome the intuitions of System 1. By admitting that System 1 can guide System 2, we have another reason to question informational encapsulation.

## 6. Conclusion

Altogether, this paper is meant to be another brick in the wall of questioning dichotomous interpretations. I think I have shown that a dichotomous interpretation of a two-system approach leads to quite unattractive consequences. First, a dichotomous distinction of properties describing cognitive processes suggests that each property stands in an either–or relation to its counterpart and thereby ignores the possibility of graduated properties. Second, by assuming that the bundles of properties of each system necessarily co-occur with each other, it is impossible to account for other potential combinations of properties for which we do have empirical evidence. Last but not least, the radical claims about inaccessibility do not hold in principle for all implicit cognitive processes, since there are several examples of how distinct processes can mutually influence each other.

Therefore, I admonish: beware of dichotomous conceptions! Instead of rejecting the idea of a two-system framework as a whole, I suggest considering a less dichotomous interpretation of it, which can account for the fact that we are not always at the mercy of automatic operations!

Future work should elaborate a framework for how to conceptualize in-between cases which are not captured by a dichotomous framework specifying only extreme cases. Promising suggestions have come already from psychologists and philosophers, suggestions which aim to describe a gradual development of the properties of cognitive processes. For instance, Dienes and Perner (1999) outline a number of levels of

implicitness, with the final level appealing to notions of consciousness. Karmiloff-Smith (1992) claims that many cognitive abilities may develop initially in an implicit form and then become increasingly explicit. We can make use of the two-system idea to characterize extreme cases in a continuum and still be open to other instances in this broad spectrum. Taking such a continuum approach allows us to acknowledge that the properties of both systems can change gradually and can be combined in various ways. Moreover, we can account for mutual influences of diverse processes.

## Disclosure statement

No potential conflict of interest was reported by the author.

## Notes on contributor

*Anna Strasser* received her Ph.D. at the Albert-Ludwigs-Universität Freiburg in 2004. She was an Assistant professor at Center for Cognitive Science, Albert-Ludwigs-Universität Freiburg, and a Postdoctoral Researcher & Scientific Coordinator, Berlin School of Mind and Brain / Institute of Philosophy, Humboldt-Universität zu Berlin. Since she has been a Visiting Fellow of Daniel Dennett at Tufts University, Center for Cognitive Studies in 2018, she works as a freelancing philosopher.

## ORCID

Anna Strasser http://orcid.org/0000-0002-3547-9790

## References

Baillargeon, R., Scott, R., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91*(4), 612–625. https://doi.org/10.1037/0022-3514.91.4.612

Balconi, M., Fronda, G., Venturella, I., & Crivelli, D. (2017). Conscious, pre-conscious and unconscious mechanisms in emotional behaviour. Some applications to the mindfulness approach with wearable devices. *Applied Sciences*, *7*(12), 1280. https://doi.org/10.3390/app7121280

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Erlbaum.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479. https://doi.org/10.1037/0003-066X.54.7.462

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, *391* (6669), 756. https://doi.org/10.1038/35784

Bratman, M. (2014). *Shared agency: A planning theory of acting together*. Oxford University Press.
Broadbent, D. (1958). *Perception and communication*. Pergamon Press.
Brooks, R. (1991). Intelligence without reason. *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (Vol. 1, pp. 569–595).
Brownell, C. A. (2011). Early developments in joint action. *Review of Philosophy and Psychology*, *2*(2), 193–211. https://doi.org/10.1007/s13164-011-0056-1
Buckner, C., & Fridland, E. (2017). What is cognition? Angsty monism, permissive pluralism(s), and the future of cognitive science. *Synthese*, *194*(11), 4191. https://doi.org/10.1007/s11229-017-1505-x
Butterfill, S., & Apperly, I. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036
Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, *28*(2), 141–172. https://doi.org/10.1111/mila.12014
Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229
Davidson, D. (1980). *Essays on actions and events*. Oxford University Press.
Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*(5), 735–808. https://doi.org/10.1017/S0140525X99002186
Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629
Evans, J., & Stanovich, K. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–224. https://doi.org/10.1177/1745691612460685
Fodor, J. (1983). *The modularity of mind*. MIT Press.
Fodor, J. (1987). *Psychosemantics*. MIT Press.
Frank, M. (2015). *Präreflexive Subjektivität. Vier Vorlesungen*. Reclam.
Gallagher, S., & Zahavi, D. (2007). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. Routledge.
Gallagher, S., & Zahavi, D. (2008). *Phenomenological approaches to self-consciousness*. Stanford Encyclopedia of Philosophy.
Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692
Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*, 59–127. https://doi.org/10.1016/B978-0-12-385522-0.00002-0
Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008a). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, *24*(4), 218–225. https://doi.org/10.1027/1015-5759.24.4.218
Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008b). When "Just Say No" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, *44*(2), 370–377. https://doi.org/10.1016/j.jesp.2006.12.004
Gopnik, A. (2003). The theory theory as an alternative to the innateness hypothesis. In L. M. Antony (Ed.), *Chomsky and his critics* (pp. 238–254). Blackwell.
Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. MIT Press.

Hutto, D., Herschbach, M., & Southgate, S. (2011). Mindreading and its alternatives. *Review of Philosophy and Psychology*, *2*(3), 375–395. https://doi.org/10.1007/s13164-011-0073-0

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

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology." Human Perception and Performance*, *9*(4), 497–509. https://doi.org/10.1037/0096-1523.9.4.497

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press/Bradford Books.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, A. S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*(5), 871–888. https://doi.org/10.1037/0022-3514.78.5.871

Kovács, Á., Téglás, E., & Endress, A. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Leslie, A., Friedman, O., & German, T. (2004). Core mechanisms in theory of mind. *Trends in Cognitive Sciences*, *8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Levin, D., & Banaji, M. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology. General*, *135*(4), 501–512. https://doi.org/10.1037/0096-3445.135.4.501

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology*, *39*(2), 367–386. https://doi.org/10.1037/h0080066

Lurz, R. (2011). *Mindreading animals: The debate over what animals know about other minds*. MIT Press.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(1), 126–135. https://doi.org/10.1037//0278-7393.14.1.126

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *391*, 756. https://doi.org/10.1038/264746a0

Michael, J., Sebanz, N., & Knoblich, G. (2016). The sense of commitment: A minimal approach. *Frontiers in Psychology*, *6*, 1968. https://doi.org/10.3389/fpsyg.2015.01968

Nier, J. A. (2005). How dissociated are implicit and explicit racial attitudes? A bogus pipeline approach. *Group Processes & Intergroup Relations*, *8*(1), 39–52. https://doi.org/10.1177/1368430205048615

Pacherie, E. (2013). Intentional joint agency: Shared intention lite. *Synthese*, *190*(10), 1817–1839. https://doi.org/10.1007/s11229-013-0263-7

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions through eye gaze. *Proceedings of the National Academy of Sciences*, *112*(13), 4170–4175. https://doi.org/10.1073/pnas.1415250112

Peng, C., Henry, I., Mietus, J., Hausdorff, J., Khalsa, G., Benson, H., & Goldberger, A. (2004). Heart rate dynamics during three forms of meditation. *International Journal of Cardiology*, *95*(1), 19–27. https://doi.org/10.1016/j.ijcard.2003.02.006

Peng, C., Mietus, J., Liu, Y., Khalsa, G., Douglas, P., Benson, H., & Goldberger, A. (1999). Exaggerated heart rate oscillations during two meditation techniques. *International Journal of Cardiology*, *70*(2), 101–107. https://doi.org/10.1016/S0167-5273(99)00066-2

Quirk, G. J., & Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology*, *33*(1), 56–72. https://doi.org/10.1038/sj.npp.1301555

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a

dual-task study of adults. *Cognition*, *117*(2), 230–236. https://doi.org/10.1016/j.cognition.2010.08.003

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology*, *44*(2), 386–396. https://doi.org/10.1016/j.jesp.2006.12.008

Robbins, P. (2017). *Modularity of mind* (E. N. Zalta, Ed.) The Stanford Encyclopedia of Philosophy. Retrieved August 01, 2018, from https://plato.stanford.edu/archives/win2017/entries/modularity-mind

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Sassenberg, K., & Moskowitz, G. B. (2005). Don't stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology*, *41*(5), 506–514. https://doi.org/10.1016/j.jesp.2004.10.002

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*(6814), 788. https://doi.org/10.1038/35048669

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (2nd ed., pp. 739–811). Wiley.

Smith, C. T., & Nosek, B. A. (2011). Affective focus increases the concordance between implicit and explicit attitudes. *Social Psychology*, *42*(4), 300–313. https://doi.org/10.1027/1864-9335/a000072

Starr, A., & Phillips, L. (1970). Verbal and motor memory in the amnestic syndrome. *Neuropsychologia*, *8*(1), 75–88. https://doi.org/10.1016/0028-3932(70)90027-8

Stewart, T., Latu, I., Kawakami, K., & Myers, A. (2010). Consider the situation: Reducing automatic stereotyping through situational attribution training. *Journal of Experimental Social Psychology*, *46*(1), 221–225. https://doi.org/10.1016/j.jesp.2009.09.004

Strasser, A. (2018). Social cognition and artificial agents. In V. Müller (Ed.), *Philosophy and theory of artificial intelligence 2017* (pp. 106–114) SAPERE. Springer.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643–662. 6 doi:10.1037/h0054651

Thagard, P. (2019). *Cognitive science*. The Stanford Encyclopedia of Philosophy (Spring 2019 Ed), E. N. Zalta (Ed.), https://plato.stanford.edu/archives/spr2019/entries/cognitive-science/

Töllner, T., Müller, H. J., & Zehetleitner, M. (2012). Top-down dimensional weight set determines the capture of visual attention: Evidence from the PCN component. *Cerebral Cortex*, *22*(7), 1554–1563. https://doi.org/10.1093/cercor/bhr231

Vesper, C., Butterfill, S., Sebanz, N., & Knoblich, G. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8/9), 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development*, *77*(3), 640–663. https://doi.org/10.1111/j.1467-8624.2006.00895.x

Wojnowicz, M., Ferguson, M. J., Dale, R., & Spivey, M. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*(11), 1428–1435. https://doi.org/10.1111/j.1467-9280.2009.02448.x