Consciousness and inference to the best explanation: Compiling empirical evidence supporting the access-phenomenal distinction and the overflow hypothesis

Asger Kirkeby-Hinrup\textsuperscript{a,b,*}, Peter Fazekas\textsuperscript{b,c}

\textsuperscript{a} Department of Philosophy and Cognitive Science, Lund University, Sweden
\textsuperscript{b} Cognitive Neuroscience Research Unit, CFIN, Aarhus University, Denmark
\textsuperscript{c} Centre for Philosophical Psychology, University of Antwerp, Belgium

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Theories of consciousness
Empirical evidence
Inference to the best explanation
IBE
Overflow
Access consciousness
Phenomenal consciousness

\textbf{ABSTRACT}

A tacit assumption in the field of consciousness studies is that the more empirical evidence a theory can explain, the better it fares when weighed against competitors. If one wants to take seriously the potential for empirical evidence to move forward debates in consciousness studies, there is a need to gather, organize, validate, and compare evidence. We present an inference to the best explanation (IBE) process on the basis of empirical support that is applicable in debates between competing theories of consciousness. Our proposed IBE process consists in four steps: Assimilate, Compile, Validate, and Compare. Until now, the vast majority of the work in the field has consisted in gathering empirical evidence for theories i.e., the assimilation step. To illustrate the feasibility of our proposed IBE process, and what it may look like when applied in practice, we deliver a complete collection (the compilation step) of empirical support for the distinction between A-Consciousness and P-Consciousness and the overflow hypothesis. Finally, we offer an example of the validation step, by scrutinizing the interpretation of aphantasics’ performance on retro-cue paradigms offered in the literature in support of the overflow hypothesis. The compilation we deliver here is the first effort in the IBE process, the end result of which — hopefully — will be the ability of the research community to carry out side-by-side comparisons of theories and the empirical phenomena they claim to explain, i.e., the comparison step.

1. Introduction

Through time, philosophy has offered numerous different ways to conceive of consciousness. Over the last half century or so, a handful of different lines of thinking have crystalized and clusters of philosophers and empirical scientists each advancing competing lines of thought have emerged. Pertinent to this is the rapid development of the relevant empirical sciences (e.g. neuroscience and cognitive science) over the last decades which has facilitated the development and comparison of theories of consciousness from a new, viz. empirical, perspective. The practice of leveraging support from the empirical domain is becoming more prevalent in debates between competing theories of consciousness (Block, 2005; 2007; Lau & Brown, 2019; Lau & Rosenthal, 2011; Weisberg, 2014). This also includes contesting claims of empirical support or debunking criticism based in empirical findings (e.g. Odegaard, Knight and Lau, 2016).

\* Corresponding author at: Center for Functionally Integrative Neuroscience, Norrebrogade 44, Building 1A, 8000 Aarhus C, Denmark.
\textit{E-mail addresses:} asger.kirkeby-hinrup@fil.lu.se, akh@cfin.au.dk (A. Kirkeby-Hinrup).

https://doi.org/10.1016/j.concog.2021.103173
Received 1 March 2021; Received in revised form 18 July 2021; Accepted 19 July 2021
2017; Malach, 2011; Brinck & Kirkeby-Hinrup, 2017; Kirkeby-Hinrup, 2014, 2016, 2020), and arguing against philosophical claims on empirical grounds (e.g. Sebastián, 2014). This heightened and increasing focus on empirical corroboration of theories of consciousness has been termed the empirical turn in the philosophy of mind.

The empirical turn partly is driven by the fact that many debates in consciousness studies are fueled by fundamentally different conceptions of consciousness, i.e., disagreement about how to even conceive of the central explanandum. However, one thing that participants in the debates mostly do agree upon is that consciousness can be naturalized, which usually is taken to mean that consciousness depends on activity in the brain. Therefore, the idea that empirical evidence can play a special role in moving forward the debates enjoys some consensus between researchers who disagree on almost everything else.

Thus, the rationale for applying empirical data in debates on consciousness is the belief that empirical evidence may substantiate thought experiments, increase the plausibility of conceptual claims by confirming their agreement with empirical findings, choose between hypotheses that are equivalent from a conceptual point of view, and much more.

Behind the empirical turn lies the shared assumption that the more empirical evidence a particular theory can explain, the stronger it stands against competitors. Given the empirical turn, if you can show that a theory of consciousness can account for a wider range of empirical data than its competitors, you can make the case that it is the best explanation overall. This sentiment is shared by Ned Block, who is central to what follows below. Block writes (2007, p. 486):

I have in mind […] the familiar default ‘method’ of inference to the best explanation, that is, the approach of looking for the framework that makes the most sense of all the data […]

The purpose with inferring to the best explanation in this way is, according to Block, to identify the theory of consciousness (here called ‘framework’) that can make the most sense of all the data. On this conception, inference to the best explanation implies evaluating and comparing theories of consciousness on the bases of their predictive and explanatory powers in the empirical domain.

Occasionally, inferring to the best explanation is considered a suboptimal solution to problems that are not solvable through logical deduction, or elimination of hypotheses through falsification (see e.g., Minnameier, 2010). Many seem to fear that shortcomings such as these also pertain to central problems in consciousness studies. For instance, Cohen and Dennett have argued (2011) that some theories in the field are unfalsifiable. Similarly, Jacob Hohwy has noted that (at least) “… the current NCC methodologies taken separately most likely cannot gain much direct evidence for or against [content focused theories of consciousness].” (Hohwy, 2009, p. 435). To a large extent, it of course depends on the specific formulation of each hypothesis about consciousness whether it can be developed to the point where attempting falsification is feasible. Most likely, one or more competing hypotheses will eventually be discarded through falsification-type experiments in the future. However – and this is the critical part — unless every hypothesis is amenable to falsification experiments, inference to the best explanation is necessary at some point. Here, we will set aside this issue, and proceed from what the present holds, namely that the consensus of current debates in the interdisciplinary field of consciousness studies is the assumption that the hypothesis that meshes best with empirical findings is preferable.

2. Assimilate, compile, validate and compare

Taking seriously the project of deploying empirical evidence to infer to the best explanation in consciousness studies (which we will call the IBE project) involves time consuming and unglamorous work. Nevertheless, for reasons sketched above, and further explained below, inferring to the best explanation appears to be the only viable process through which to move forward the debates. Crucially, in our view, even the preliminary work leading up to the actual comparison of theories on the bases of their empirical merits, will serve to stimulate, bring into focus, and move forward the current debates. It is also worth noting that classical notions of inference to the best explanation do not straightforwardly apply to the debates in consciousness studies. For instance, in some classical literature on inference to the best explanation (Harman, 1965; see also explication in Campos, 2011), every intermediate step (lemma) must be true, to not undermine the final inference. This need not necessarily be the case in here, because it is possible that one or more pieces of empirical support may be rejected without this undermining the whole IBE process. Finally, in addition to these reasons, because a concrete and feasible process to follow in the domain of empirically minded consciousness studies is clearly preferable, we here operate with a rational reconstruction of the process of inferring to the best explanation when it is applied to competing theories of consciousness based on their respective empirical support. The process consists in four discrete steps, which we will turn to next.

2.1. Step one: assimilate

The first step consists in the proponents of each theory assimilating empirical evidence they think supports their theory. Assimilation, by and large, makes up the vast majority of work carried out by participants in the debate until now. To illustrate the extent of work on assimilation within just one theory, we refer to the compilation below. As suggested in the introduction, the focus on assimilation of empirical support across the field is indicative of the consensus that empirical evidence has a special role in the debate, which is the premise of the IBE project.

1 To be clear, this is not to diminish the value of philosophy of science or epistemology. Certainly, both fields have important insights that are relevant to our endeavors in consciousness studies.

2 It is worth noting, of course, that the compilation and the validation steps could be done in reverse order, in the sense that one could validate each piece of purported empirical support before adding it to the compilation.
2.2. Step two: compile

The second step consists in compiling every piece of empirical support claimed by each theory. Thus, compilation means collecting, logging, and organizing all the empirical evidence proposed in each assimilation process. So far, this aspect of the IBE project has been largely ignored. There have been sporadic attempts in this vein in the form of review articles (e.g. Block, 2007; Dehaene et al., 2017; Doerig et al., 2021; Lau and Rosenthal, 2011, and more recently Yaron et al. 2021). However, there has been no ongoing, systematic, and exhaustive effort regarding the compilation aspect of the IBE project. While informative and relevant, reviews alone do not suffice for compilation, since they tend to focus on only the newest, strongest or most fashionable empirical findings. Presumably, there are reasons for the lack of attention to compilation. For instance, the pressure in academia to publish novelties contra the time-consuming nature of reading, identifying, distilling, and compiling claims of empirical support already published. However, the compilation aspect is crucial if we want to take seriously the IBE project. Put bluntly, if we want to compare theories based on their empirical support, we need to know what their respective empirical support is. That much seems obvious.

2.3. Step three: validate

The third step in inferring to the best explanation on the bases of empirical support consists in assessing whether the evidence in question can indeed provide the support claimed. The aim here is validating, correcting, or rejecting evidence on a case-by-case basis. A precondition to be able to arbitrate between competing theories of consciousness on the bases of their empirical support is independent assessment of the proposed empirical evidence, and the arguments through which it supposedly supports a theory of consciousness. This is the validation step. Validation has counterparts in the notion of replication in empirical sciences and fact checking in public discourse. Like its extra-disciplinary counterparts, independent assessment of proposed empirical evidence has received relatively little attention. However, such work has a crucial role to play in empirically minded philosophy (as in interdisciplinary research generally) in clarifying the complex relation that holds between empirical and philosophical claims e.g., by pointing to reasoning errors, implausible interpretations and errors that arise from cognitive biases.

There are at least two elements to this. The first is to assess whether the suggested interpretation of the empirical results is plausible. This is particularly important when more than one empirical finding or phenomenon is combined to make an argument. The second element consists in assessing the connections between a given interpretation of the empirical data and a theoretical claim. By theoretical claims, we here mean claims cached in the conceptual framework of a particular theory of consciousness. These claims are often easy to spot because they are likely to deploy concepts that are proprietary, as it were, to a given theory (e.g. “global broadcasting”, “overflow” etc.). By ‘assessing the connections’ we mean investigating how the interpretation features in an argument purporting to arrive at a particular conclusion in the context of a given theoretical claim. Like the compilation step of the IBE project discussed above, the process of validating claims of empirical support has not received much attention (but see e.g., references in the introduction above for a few examples of this work, where claims of empirical support are debunked). Again, there are likely uncontroversial socio-scientific reasons for this. For instance, it is hard to publish ‘negative’ results, in the sense that just pointing out that some specific empirical claim in a very broad and complex debate is wrong is not of huge interest to a wide readership, and therefore not attractive content for publishers. Another plausible reason is that most participants in the debates primarily have been concerned with bolstering their preferred theory with empirical support (i.e., engaged with the Assimilation step). That this is the case is indicated by the fact that most opponents of overflow have been more concerned with giving no-overflow interpretations of empirical evidence leveraged by proponents of overflow, than with debunking the overflow interpretation of said evidence. Thus, less time has gone into scrutinizing support that other theories have claimed (and people not engaged in the debates have little incentive to dedicate precious time to this).

In any case, the process of validation serves a critical purpose in inferring to the best explanation since, clearly, we should not count an empirical finding among the phenomena that supports a theory, if it does not in fact support the theory.

2.4. Step four: compare

With the empirical support of each theory compiled and validated, the fourth and final step is the comparison of the theories. Initially, this may sound straightforward, but there are several outstanding questions pertaining to this step as well. We will not go in depth with these here, but merely note that some weighing of the distinct pieces of empirical support seems reasonable. For instance, one might argue that if a theory can explain widely replicated empirical findings of very prevalent features of consciousness in neurotypical adults, this seemingly should weigh more heavily in the comparison than an interpretation of rare phenomena based on few studies in neuropathological subjects. Such weights may be modelled using varieties of confirmation theory (e.g. Laymon, 1987; Christensen, 1999). In any case, exactly to carry out the comparison we will leave open for now.

3. Motivating the IBE project

Debates between competing theories of consciousness have been an ongoing affair in the conceptual domain over the last several decades. With the empirical turn these debates have increasingly turned to the empirical domain. However, what we have seen in the debates so far is that when proponents of a theory offer an interpretation of an empirical phenomenon, the interpretation is conditional on acceptance of the conceptual framework of the interpreter but is otherwise coherent and consistent. Thus, it seems that the disagreements from the conceptual domain have merely migrated to the empirical domain, which is deeply problematic given that these disagreements were exactly what the application of empirical evidence was supposed to help us dissolve. Furthermore, it makes...
comparisons of competing interpretations difficult to say the least, since each interpretation will usually be coherent if only one accepts certain conceptual prerequisites. It is in response to these issues we propose the IBE project. Our proposal derives from the observation that another approach seems necessary if we want to make progress in determining which theory meshes the best with empirical data. We acknowledge that in addition to what we have proposed here, there is still much theoretical work to be done to develop and refine the IBE process. This is especially true regarding the comparison step, as we noted above. However, while we have left open how exactly to go about this, we feel confident that one or more of the available methods from the philosophy of science, whether it be Bayesian estimation, formal confirmation theory, or something else, will be suitable to carry out the comparison. In sum, we readily acknowledge that much hinges on how exactly comparison will eventually be carried out, and the IBE project in general will benefit greatly from the future involvement of philosophers of science.

4. The IBE project in practice

Above, we introduced a rational reconstruction of the IBE project in debates between competing theories of consciousness and argued why it constitutes a path forward for interdisciplinary consciousness studies. With this in place, we will, in the rest of this article, support this with practical examples of the work involved. In the next two sections we deliver first a complete (at the time of writing, and to the best of our knowledge) compilation of the empirical support of one theory on offer in the debate and second, examples of what validation would look like in this context. The compilation offers the most comprehensive compilation to date of empirical support claimed in favor of the theory proposed by Ned Block. Block himself has earlier categorized this as a ‘biological theory’ (Block, 2009, p. 1113). In philosophical terms, the theory is sometimes characterized as a first-order theory (see e.g., Lau and Rosenthal, 2011, p. 365). Given the close association with the recurrent processing theory of Victor Lamme (e.g. 2003), it is sometimes called (self-)reflexive first-order theory. The reason it is considered a ‘first-order’ theory, is that a first-order mental state, such as a visual state, does not need any external state or process (e.g. a higher-order state, or global broadcasting) to become phenomenally conscious. The ‘reflexive’ moniker derives from the connection to Lamme’s recurrency theory, according to which recurrent feedback is necessary for phenomenal consciousness. Recurrent feedback can reasonably be conceived of as a mental state having a reflexive relation to itself. Be that as it may, there are two central claims to Blocks position, that we here will take to encapsulate Block’s theory regardless of what monikers are ascribed to it. The first is what we will term the A-P distinction. The A-P distinction distinguishes between phenomenal consciousness (P-consciousness) and access consciousness (A-Consciousness). Starting with the former, P-consciousness consists in the peculiar first-person experience of being conscious. This makes P-consciousness essentially experiential by virtue of its reliance on the immediate subjective experience of being in a mental state, e.g. feeling the pangs of jealousy or perceiving the redness of a tomato. A-consciousness, on the other hand, is essentially functional in the sense that a mental state is A-conscious if it is poised to figure in cognitive functions such as inferences or report. In short, a state is A-conscious if it can figure in rational control and report, and it is P-conscious if a subject experiences being in the state. For Block then, A-consciousness drives the cognitive domain, while P-consciousness drives the experiential domain. The distinction is not merely a conceptual claim useful for analysis, but rather the (more substantial) empirical claim that the two types of consciousness are differentially instantiated in the brain and may come apart in the sense that A-consciousness may occur without P-consciousness and vice versa. The sharp distinction between the cognitive and the phenomenal domain challenges theorists to distinguish A- from P-consciousness empirically. This is what Block has called the methodological puzzle of consciousness (Block, 2007). The puzzle consists in the fact that methods for probing P-consciousness seemingly must “go through” A-consciousness, which threatens a confound.

The second crucial claim we take to be encapsulated by Block’s theory is the overflow hypothesis (we will use ‘overflow’ for brevity). Block’s overflow claim posits that the content of P-consciousness harbors more information than a subject can access. So, subjects can have conscious experiences they are not — and cannot be — aware of having. Or put more simply: because the phenomenal domain outstrips the cognitive domain, we experience more than we can access. The overflow claim is sometimes stated in stronger terms, for instance when Block says: “I will assume it is meaningful to suppose that GK has an experience that he does not and cannot know about” (Block, 2008, p.292). How exactly to understand this claim is underdetermined. One way is that all P-conscious state-types are contingently available for access. On this understanding the overflow claim is the result of an informational bottleneck: accessing a given P-conscious state token X prevents access to another token Y. Only one, but not both states, can pass through the bottleneck—and thus become A-conscious. An alternative way of understanding this is by stripping the contingency and holding that some P-conscious states are essentially or necessarily un-accessible.

3 We are grateful to an anonymous referee for prompting us to underscore this point.
4 It is worth noting, however, that by and large philosophers of science are conspicuously absent from the interdisciplinary field of consciousness studies in general. We speculate that at least three factors may be contributing causes to this absence. The first pertains to the interdisciplinary nature of consciousness studies, which requires a philosopher of science to spend significant time to build up expertise in other disciplines (e.g. philosophy of mind and the relevant empirical disciplines). The second may be that workable datasets suitable for the deployment of philosophy of science have been scarce, thus giving philosophers of science little to work with. The third possible cause is simply a lack of awareness that the field of interdisciplinary consciousness studies both is an apt subject of study and in need of their expertise. This paper (and our rational reconstruction of the IBE project) goes some way to address the first two of these possible causes, by introducing a large dataset and the relevant theoretical context from the philosophy of mind. In separate work, one of the authors has addressed the third possible cause (Kirkby-Hinrup, in preparation).
5 We should note that we will here deploy the term A-consciousness, even if this term in recent years has given way to the term ‘cognitive access’ (Fazekas and Overgaard, 2018).
5. Compilation

In this section, we present the compilation of empirical evidence proposed in favor of the A-P distinction and overflow in the literature. Each of the empirical phenomena in the compilation should be seen in the light of these two claims. All entries in the compilation argue directly or indirectly for either the A-P distinction or Overflow.

Since this is all done in the context of the IBE project, what we are counting are empirical phenomena that have been offered an explanation in light of the A-P distinction or Overflow. This means we are not counting individual paradigms or replications. However, in some cases, there is a great deal of variance in the paradigms used to investigate an empirical phenomenon and interpretations of the results, and this we attempt to reflect in our summary. We remain neutral on the relative evidential weight (see above) of the phenomena. Finally, we aim to remain strictly within the compilation stage of the IBE process. This means that we also remain neutral on whether the arguments provided to connect an empirical phenomenon with the claims are sound or convincing. The aim presently is to present soberly the evidence and its interpretation and — where necessary — neutrally reconstruct the underlying assumptions, inferences, or arguments. In the same vein, we do not include objections to the interpretations or the arguments in which they figure. Similarly, we are not including competing interpretations, since a given alternative interpretation in light of a competing theory X — according to the way we conceive of the IBE project — rightfully belongs in the compilation of empirical evidence of theory X. This, of course, does not entail that all explanations are equally good, or indeed sound. However, objections to and comparisons of interpretations belong respectively in the validation and comparison stages of the IBE project.

Many of the entries in this compilation are cached in terms so-called proxy debates. By proxy debates, we here mean debates about particular phenomena and their putative implications for which proponents of the A-P distinction have defended a specific point. To qualify as a proxy debate in the context of this paper, the viewpoint in question should be an extension of — or have important implications for — the A-P distinction and/or overflow. The rate of occurrence of such proxy debates has been steady over the last couple of decades; early examples of these include e.g. debates concerning unconscious perception and non-conceptual content. More recent examples are debates over whether perception is rich or sparse, and perceptual precision. One upshot of this is that many of the entries will reference parts (such as terminology) of the particular proxy debate in which they were advanced. In our exposition of each entry, we do not delve deep into the details of each of these debates, but only succinctly note the context of the proxy-debate and how the interpretation in light of this context applies to the A-P distinction or overflow. This does not entail that the proxy debates and their accompanying terminology is irrelevant of course. However, scrutiny and assessment of this belongs in the validation step for each entry, and not in the compilation.

Before we proceed with the compilation, there is a similar issue concerning the relationship between consciousness and attention, that is relevant to address in this preamble. Several of the entries below depend on specific views of the relation between consciousness and attention (e.g. that attention is not necessary for P-consciousness). In this paper we do not endorse any particular view regarding the relationship between attention and consciousness, which is a hotly debated question in the literature (see e.g. Koch and Tsuchiya, 2007; Kentridge, 2011; Cohen et al., 2012; Montemayor and Haladjian, 2015; Pitts et al., 2018; Tallon-Baudry et al., 2018). Thus, in the presentation of the entries below, we are exclusively summarizing the interpretations as given by the original proponents. Thus, we wish to underscore that we remain neutral on the role of attention in relation to consciousness, and the views found below are inherited from the original authors endorsing a given claim. Just like with the proxy debates mentioned above, scrutiny of such auxiliary conceptual and/or theoretical commitments and the work they do in the interpretation of a given phenomenon rightfully belongs in the validation stage for said phenomenon.

We have divided the phenomena loosely into three categories, depending on what the phenomenon in question is deployed in support of, or to argue for. The first two categories concern the A-P distinction, and collect empirical phenomena invoked in relation to the A-P distinction. The difference between the two categories is that the latter contains entries specifically addressing the putative neural underpinnings of the A-P distinction, whereas the first contains all other entries (mainly, but not exclusively, from cognitive science and psychology). The third and final category collects all entries deployed in support of overflow. Due to the interwoven nature of decades; early examples of these include e.g. debates concerning unconscious perception and non-conceptual content. More recent examples are debates over whether perception is rich or sparse, and perceptual precision. One upshot of this is that many of the entries will reference parts (such as terminology) of the particular proxy debate in which they were advanced. In our exposition of each entry, we do not delve deep into the details of each of these debates, but only succinctly note the context of the proxy-debate and how the interpretation in light of this context applies to the A-P distinction or overflow. This does not entail that the proxy debates and their accompanying terminology is irrelevant of course. However, scrutiny and assessment of this belongs in the validation step for each entry, and not in the compilation.

Before we proceed with the compilation, there is a similar issue concerning the relationship between consciousness and attention, that is relevant to address in this preamble. Several of the entries below depend on specific views of the relation between consciousness and attention (e.g. that attention is not necessary for P-consciousness). In this paper we do not endorse any particular view regarding the relationship between attention and consciousness, which is a hotly debated question in the literature (see e.g. Koch and Tsuchiya, 2007; Kentridge, 2011; Cohen et al., 2012; Montemayor and Haladjian, 2015; Pitts et al., 2018; Tallon-Baudry et al., 2018). Thus, in the presentation of the entries below, we are exclusively summarizing the interpretations as given by the original proponents. Thus, we wish to underscore that we remain neutral on the role of attention in relation to consciousness, and the views found below are inherited from the original authors endorsing a given claim. Just like with the proxy debates mentioned above, scrutiny of such auxiliary conceptual and/or theoretical commitments and the work they do in the interpretation of a given phenomenon rightfully belongs in the validation stage for said phenomenon.

We have divided the phenomena loosely into three categories, depending on what the phenomenon in question is deployed in support of, or to argue for. The first two categories concern the A-P distinction, and collect empirical phenomena invoked in relation to the A-P distinction. The difference between the two categories is that the latter contains entries specifically addressing the putative neural underpinnings of the A-P distinction, whereas the first contains all other entries (mainly, but not exclusively, from cognitive science and psychology). The third and final category collects all entries deployed in support of overflow. Due to the interwoven nature of decades; early examples of these include e.g. debates concerning unconscious perception and non-conceptual content. More recent examples are debates over whether perception is rich or sparse, and perceptual precision. One upshot of this is that many of the entries will reference parts (such as terminology) of the particular proxy debate in which they were advanced. In our exposition of each entry, we do not delve deep into the details of each of these debates, but only succinctly note the context of the proxy-debate and how the interpretation in light of this context applies to the A-P distinction or overflow. This does not entail that the proxy debates and their accompanying terminology is irrelevant of course. However, scrutiny and assessment of this belongs in the validation step for each entry, and not in the compilation.

6 One reviewer expressed concerns that compiling evidence for one theory at a time would result in a one-sided picture, and that the text would be “unbalanced” as a review. We urge readers to keep in mind that the aim of compiling evidence is not to advance any theory of consciousness, and we are confident that readers appreciate that the context of the compilation is the IBE project. Furthermore, we here again want to underscore that a compilation for every theory is necessary, and preparations for the next ones are ongoing. The same reviewer highlighted the idea that if two theories can account equally well for a given piece of empirical data, these would “cancel out” each other, and the data could not be counted in support of either theory. For this reason, the reviewer suggested, it would not make sense to compile data separately. We disagree with this objection. Firstly, there are a handful of major groups of theories, each of which contains various proposals of specific theories. Not all theories have considered all phenomena as of yet, and there is still some way to that point, therefore currently no phenomenon is actually (fully) cancelled out. Furthermore, because novel theories or variations of existing ones are likely to appear in the future, this is an open-ended endeavor with no non-arbitrary cut-off for when we should stop counting a phenomenon. Thus, it makes sense to keep track of each theory separately, since this allows us to diagnose which phenomena newly developed theories have yet to account for, which would not be possible if we stopped keeping track of phenomena that were cancelled out. To boot, there is the possibility of future counter evidence or counter arguments to a given theory’s interpretation of a phenomenon may appear, which would “cancel a cancellation” as it were. By having compilation, validation and comparison in separate stages, such alterations in the datasets pose no problems. Finally, having the compilations separately available caters to future researchers and may thus facilitate new research, counter arguments, or developments, something that would not happen if we stopped keeping track of phenomena we thought were “cancelled out”.

7 We are thankful to an anonymous referee for prompting an elaboration of these issues.
of the debates, it is worth bearing in mind that some phenomena could rightfully belong in more than one of these categories.

Within each of the three categories we have ordered loosely the entries according to the rough amount of attention each has received. Thus, we have tried to put the most widely discussed phenomena first, and the more esoteric toward the end of each section. Importantly, the reader is to not infer anything beyond a ‘headliners first’ approach from this ordering.

6. A-P distinction: Dissociation at the psychological level

Blindsight

The Blindsight phenomenon derives from subjects who, due to damage to the primary visual cortex (V1), have ‘blind’ areas in their visual fields. When asked about their experiences of stimuli presented in the blind areas of the visual field, the subjects say they did not experience anything. Consequently, they deny being able to identify stimuli presented to the blind areas of the visual field. Nevertheless, when prompted to guess anyway, some subjects perform reliably on certain features. Some can guess above chance between a limited set of simple forms, do limited color discriminations and can make appropriate grasping motor actions towards objects located in the blind areas. On the basis of this discrepancy between what subjects report experiencing and their actual performance, Block (1995) argues that blindsight is an example of A-consciousness without P-consciousness. A similar albeit less explicit interpretation is proposed by Austen Clark (2010). In another paper, Block (2005) references an experiment using blindsighted monkeys (Cowey and Stoerig, 1997), who learned a visual discrimination on their intact part of the visual field, and subsequently were able to perform the same visual discrimination, when the task was presented in their blind areas. Supposedly this indicates the dissociation between P-consciousness and A-consciousness in the monkeys.

Perceptual organization

Perceptual organization (sometimes called perceptual integration or binding) is the ability of the visual system to “integrate and bind information across spatially separated sets of neurons to infer perceptual […] properties” (Fahrenfort and Lamme, 2012, p. 139). This is explained by Fahrenfort and Lamme (2012) in terms of the A-P distinction (theories they collect under the moniker “phenomenal-access theories”). Fahrenfort and Lamme suggest that because perceptual organization serves as input to attentional mechanisms, A-consciousness cannot explain the contents of phenomenal experience, and therefore cannot fully explain phenomenology (Fahrenfort and Lamme, 2012, p. 139). This relies on the assumption that attention is a key element of A-consciousness although they do not state this explicitly. The conclusion Fahrenfort and Lamme seek is that if A-consciousness does not fully explain phenomenology, then the A-P distinction is the preferable hypothesis.

Visual neglect

Visual neglect is a phenomenon in which subjects ignore — or fail to attend to — one side of the visual field, relative to the median line of the body (Gauthier, Dehaut and Joanette, 1989). The exact neural underpinnings of this deficit are still in dispute (see e.g., Mort et al., 2003). Despite their ignorance of content on the one side of the visual field, subjects with neglect nevertheless show effects of the ignored contents when reporting on stimuli that crosses the midline. This effect can be found when subjects are presented with variations of the Mueller-Lyer illusion and Kanizsa figures. Block (2012a) suggests that subjects partly maintain A-consciousness of the visual stimuli presented in the ignored visual hemifield, despite the appearance that they are not P-conscious of said stimuli. This interpretation is echoed by Clark (2010), who cashes out the phenomenon using a similar distinction between “conscious” and “sensory registration without awareness”, which prima facie maps well onto the A-P distinction (see also Clark, 2010, footnote 11, for terminological considerations in relation to overflow).

Dishwasher examples

‘Dishwasher examples’ are easily relatable phenomena from everyday life. Their strength is recognizability to the audience (e.g. in a classroom) and their explanation in terms of a theory, or to illustrate theoretical claims, that helpfully illuminates the theory or a claim thereof. In this sense, dishwasher examples are similar to what Daniel Dennett calls intuition pumps (Dennett, 1984). One example (Block, 1995) deployed to illustrate the difference between P-consciousness and A-consciousness (qua example of P-consciousness without A-consciousness) is the following: You are engaged in intense conversation, when suddenly at time t you realize that the noise from a deafening pneumatic drill has been coming through your window for some time. The example suggests that you were all along aware (P-conscious) of the noise, but only at time t did you become A-conscious of the noise. A variant of this deploy the noise from a refrigerator (Block, 1998); notice the refrigerator example also figures in Block’s 1999, albeit there not cast in light of the A-P distinction). These examples trade on the introspective impression that at time t you certainly had been hearing (been p-conscious of) the noise all along, even if you did not realize it (was A-conscious of it). Another classic example is the ‘absent minded driver’ of Armstrong (1968): imagine you are driving your car and absentmindedly drift into random trains of thought. At one point you realize that you have no idea about what has been happening on the road for a while. You have been on ‘autopilot’ as it were. Clearly though, representations of the road have been present, or you would have crashed the car. Block (1998) deploys a similar example of driving, but instead of absentmindedness he invokes automatic behavior during seizures.
Visual pop-out

In discussing phenomenal precision, Ned Block (2015) invokes data suggesting high-level perceptual features can exhibit visual pop-out. Because visual pop-out is thought to be a purely visual phenomenon with little to no cognitive interference, this suggests that some high-level features, e.g., faces and facial expressions, are visually represented holistically as opposed to represented as conjunctions of low-level features subsequently interpreted as faces or expressions by later processes. The upshot of this, according to Block (2015, p. 6) is that a single visual experience can contain both high-level and low-level features, independently of cognition. While this argument is made in the context of phenomenal precision, it serves as an extension of the overflow argument and A-P distinction, by way of establishing that the phenomenal contents of visual experience exceed low level features and conjunctions of these. The underlying idea is that if high-level features can drive visual pop-out, and visual pop-out is a purely perceptual phenomenon, this adds high-level properties to the non-cognitive content of which we perceptually conscious because visual pop-out seemingly is a case of “object-seeing, without object attention” (Block, 2012b).

Perceptual adaptation

Perceptual adaptation denominated occasions where a subject’s perceptual systems become biased towards perceptual features or properties. Here, “biased” means a change in sensitivity of the relevant neural elements to the feature or property in question. The neural elements can either be macro-level elements, such as brain regions, or functionally defined distributed faculties or specialized processes, or micro-level elements, such as individual neurons, e.g., edge detectors in the early visual system. In his (2015), Ned Block, develops earlier work (Block, 2014c) into an account of perceptual adaptation in which it partly depends on non-cognitive perceptual processes. He proposes that in addition to low-level properties (e.g. tilt or color) high-level properties (e.g. individual faces, or facial expressions) can also be the objects of perceptual (as opposed to cognitive) adaptation. While Block’s argument here is mainly cached in a debate with Sascha Fink (2015) about phenomenal precision, he insinuates the A-P distinction throughout. The implication is that phenomenal consciousness encompasses high-level properties, supposedly showing that we are p-consciousness of individuals or facial expressions. Here, Block (2015, p. 6) is that a single visual experience can contain both high-level and low-level features, independently of cognition. While this argument is made in the context of phenomenal precision, it serves as an extension of the overflow argument and A-P distinction, by way of establishing that the phenomenal contents of visual experience exceed low level features and conjunctions of these. The underlying idea is that if high-level features can drive visual pop-out, and visual pop-out is a purely perceptual phenomenon, this adds high-level properties to the non-cognitive content of which we perceptually conscious because visual pop-out seemingly is a case of “object-seeing, without object attention” (Block, 2012b).

Split brain

The split-brain phenomenon occurs in patients who have had a callosotomy, a procedure in which the corpus callosum is severed. Because the two brain hemispheres are wired contralaterally, and because the corpus callosum is the main information highway connecting them, each hemisphere loses insight into information acquired through its ipsilateral body half (information that becomes available to the opposite hemisphere). Despite the apparent severity of severing the corpus callosum, subjects do not report distortions or other significant changes to their conscious experiences. However, in certain experimental situations, the experimenter can control the stimulus presentation in a manner that ensures the stimulus only reaches one of the two hemispheres. In this way, it is possible to present a task to the right brain half, which the subject carries out with her left hand (that is controlled by the right hemisphere), all the while the subject verbally reports (because speech control is handled in the left hemisphere) that no task was presented, so how could she possibly comply. This effect puts pressure on the subjects’ reports that post-operatively everything appears unchanged with respect to their conscious experiences. On the one hand, the subjects’ reports suggest a picture of them as normal individuals with a unified conscious life. On the other hand, in the experimental findings, there appears to be a clear dissociation between different parts of the individual. Pinto and colleagues (Pinto et al, 2017; Pinto, De Haan, and Lamme, 2017) have offered an explanation that deploys the main tenet of the recurrent processing theory; namely that local recurrent interactions are sufficient to create consciousness (i.e., P-consciousness), even if these interactions are not part of a larger network (Pinto et al, 2017, p. 1236). This explanation accounts for the reports from the subjects about their conscious experiences. But what about the experimental findings? The split-brain data are explained by suggesting that reportability may require interhemispheric integration. Additionally, features such as verbalization and body control are also posited as requiring such integration (Pinto et al, 2017, p. 1236). Three of the main characteristics of A-consciousness are exactly reportability, conceptualization and control, and the distinction between A-consciousness and P-consciousness is tacit in their interpretation. Thus, the explanation offered for the split-brain phenomenon is that 1) subjects do have unified P-consciousness, but that 2) in certain experimental situations A-consciousness may break down and become disunified. In this explanation, 1) accounts for the subjective reports, and 2) accounts for the experimental phenomenon.

Perceptual learning

In trying to disentangle whether consciousness (understood as P-consciousness), is sufficient for perceptual leaning, or whether attention (a proxy for A-consciousness) is necessary, Meuwiese et al. (2013) deployed textured figure-ground stimuli and manipulated reportability either through masking (interfering with P-consciousness) or an inattention paradigm (interfering with A-consciousness). The conclusion offered is that attention is not necessary for perceptual learning, whereas P-consciousness is (Meuwiese et al., 2013, p. 1588). Because they found that no learning occurred for the masked stimuli, and that masking have been found to block recurrent
processing, they further conclude that recurrent processing seems to be required for perceptual learning (Meuwise et al., 2013, p. 1591), possibly by triggering NMDA-type receptors, thus mediating synaptic plasticity (Meuwise et al., 2013, p. 1592).

**Prosopagnosia**

Prosopagnosia is a neurological condition in which the subjects are unable to — or have great difficulty with — the recognition of faces. In his 1995, Ned Block deploys the capacity to ‘guess’ between two stimuli of some prosopagnosics in his work to establish the A-P distinction. While the subjects cannot recognize faces of even close relatives, they show an ability to guess above chance between familiar faces of the same occupational category (e.g. presidents: Bush and Reagan, notice also that the ability to ‘guess’ between stimuli of the prosopagnosics can be viewed as analogue to the ability of some blindsighters, see Block 1995, p. 228). Importantly, the subjects do not display covert knowledge of the identity of the faces, a facet of the data that Block holds to suggest that subjects rely on perceptual features or non-identifying yet holistic features, such as facial expressions. This upshot is the A-P distinction, since presumably the explicit recognition of faces requires A-consciousness, but when performing the guessing task, the data may suggest that prosopagnosics rely on at least some P-conscious features of the stimuli to scaffold their guesses.

**Phosphene experiences**

Phosphene experiences are flashes of light or color incurred by pressure — or magnetic stimulation — to the eyeballs (try pressing on your closed eyes with the heels of your hands). Central to the A-P distinction are differences between the two kinds of consciousness. This is sometimes done by proxy of discussing representational versus nonrepresentational content. The cardinal example of nonrepresentational content is phenomenal properties, which are assumed to be essentially linked to P-consciousness. As part of an argument for the existence of nonrepresentational content, Block (1996) proposes phosphene experiences as examples of phenomena with possibly nonrepresentational properties (see also entry on Peripheral vision below).

**Peripheral vision**

In his explanation of why in peripheral vision, phenomenologically, things may look the same, while in foveal vision they are clearly distinct, Ned Block (2014c) aims to show that P-conscious content cannot be determined by representational content. The argument goes as follows: if P-conscious content were determined by representational content, we would expect representational content in peripheral vision to be less determinate than in foveal vision. If this is not the case, P-content and representational content would come apart, because P-content exhibits this difference between foveal and peripheral precision. Evidence that at least contrast discrimination and other features do not diminish in peripheral vision, Block (2014c, p. 10-11) takes to show that representational precision does not decrease in peripheral vision. In light of the A-P distinction, phenomenal precision serves to establish that there is phenomenal content, that is not exhausted by representational (i.e. A-conscious) content. Thus, phenomenal content and representational content comes apart when comparing precision in foveal and peripheral vision, which in turn implies a distinction between A-consciousness (which supposedly relies on representational content), and P-consciousness (relying on P-content).

**Automatic behavior in petit-mal seizures**

During some types of seizures, subjects become totally unconscious, but somehow continue to exhibit goal-directed behavior. To explain this phenomenon, Block (1998, p. 28) deploys the A-P distinction. The explanation given is that while subjects are totally P-unconscious during the seizures, they nevertheless retain A-consciousness, which is sufficient to carry out the goal directed behaviors.

**Figure-ground switches**

While underscoring the distinction between A- and P-consciousness, Block (1995, p. 231) offers figure-ground switches as an example of how they interact. The idea is that different parts of the perceptual information can be accessed i.e., either through endogenous processes or by allocating attention, and this yields differences in phenomenology. Depending on what perceptual information is accessed, figure may become ground and vice-versa, and these switches affect phenomenology. Importantly, while Block suggests that attention modulates phenomenology (1995, p. 241), he maintains that P-consciousness also obtains in the absence of attention.

**Hypnotic analgesia**

Blocking a patients’ access to pain using hypnosis (hypnotic analgesia) is suggested by Block (1995) as a possible example of P-consciousness without A-consciousness. Cases where A and P-consciousness come apart are important in arguing for a A-P distinction. A key question in anesthetics is how to determine whether the subject is in fact in pain when there are no behavioral clues (including verbal reports) that suggest she is. Block (1995, p. 244) suggests that while the pain is P-conscious, the hypnosis prevents it from being A-conscious. This means the pain may be experienced, but not accessed, and explains the lack of behavioral clues.
Alexia  

Alexia is a neurological condition suggested by Block (1995) as another example of dissociation between A-consciousness and P-consciousness. In Alexia subjects have lost the ability to read words upon immediate inspection, or as Block puts it “at a glance” (Block, 1995, p. 230). The argument relies on reports from Landis et al. (1980), in which alexic subjects were able to guess words presented for time intervals too short for them to laboriously piece together their meaning (an ability retained, but limited to about one word per second, which is longer than the time the words were displayed). The fact that subjects can nevertheless guess the words allows for an interpretation of this phenomenon in terms of a dissociation between experience and access, i.e. by suggesting subjects relied on P-conscious information in their guesses.

Reverse Anton’s syndrome  

In his (1998) Block suggests what he calls ‘reverse Anton’s syndrome’ as a possible example of A-consciousness without P-consciousness. Anton’s syndrome consists in cases, where blind people do not realize they are blind. Reverse Anton’s syndrome consequently consists in individuals who do not realize that they are not blind. The example is based on a case in which an individual with extensive damage to V1 retained some visibility in the upper right quadrant of his visual field. For stimuli in this area, the subject was able to read single words, recognize faces and facial expressions (Block, 1998, p. 26). However, the individual believed himself to be completely blind, and denied having visual experiences, and when asked about how he knows what the stimuli is he says “it clicks” (ibid). There is no reason not to believe these to be sincere reports by the subject, so it seems the subject has no P-conscious visual experiences. However, the ability to identify stimuli suggests the subject is A-conscious.

Sensory substitution  

In one early paper by Block (1996), we find a precursor to the overflow position. On the first page, Block frames the debate through the question “whether there is anything in the phenomenal character of conscious experience that goes beyond the intentional, the cognitive and the functional” (Block, 1996, p. 19). And on the next page (Block, 1996, p. 20) he posits “that phenomenal character outruns representational content”. Thus, while the paper mainly concerns the relation between phenomenal qualities (qualia) and representational content, it serves as part of the foundation for the A-P distinction and overflow, and therefore warrants inclusion on this list. As part of the argument to establish that there are phenomenal qualities independent of representational content, Block deploys sensory substitution experiments by Bach-y-Rita (1971), in which blind subjects were fitted with a device that transformed visual data into tactile input, thereby allowing the subjects to “see”. The suggestion is that for the subjects in the sensory substitution experiment, it is possible to attend to — or introspect — only the experience of the tactile stimulation, independently of the informational content. Thus, there is more to sensation than representational content. This, Block suggests, also holds with respect to (normal) visual sensations (Block, 1996, p. 36). Thus, the deployment of sensory substitution serves to illustrate that the phenomenal and representational (informational) domains can be distinguished.

Shifted spectrum  

While not strictly couched in terms of the A-P distinction, Ned Block (1999) deploys an empirical argument — dubbed shifted spectrum — for the independence of phenomenal content from representational content. The shifted spectrum argument is a less extreme version of the classical inverted spectrum argument (see e.g., Shoemaker, 1982), and depends on the fact that there are individual differences (for instance based on gender, race, and age) in the peak sensitivity range of different cones in the retina. From this, Block argues that the phenomenal experiences of objective color must differ between individuals, a conclusion supported by various experimental data on color matching and color judgments (e.g. Block 1999, p. 42-43). The argument is mainly targeting representationalists, who think that phenomenal content is subsumed by representational content. However, it is in line with the A-P distinction (see e.g., Block, 1999, endnote 6), and serves to illustrate especially how to conceive of P-consciousness (qualia) as “features of experience that go beyond the experience’s representational, functional and cognitive features” (Block 1999, p. 48), and underscore that “attention and awareness must be firmly distinguished” (Block 1999, p. 64).

Unconscious response inhibition  

In arguing for the existence of unconscious perception (a proxy for the debate concerning P-consciousness without A-consciousness, which is central to establishing the A-P distinction), Ned Block (2016, p.458) points to response inhibition as an example of a phenomenon previously thought to require A-consciousness. If response inhibition can be performed without A-consciousness, the most parsimonious explanation is that they are driven by P-consciousness. Thus, it adds to the range of cognitive functions previously thought to require A-consciousness, which it turned out could be performed without A-consciousness. The goal is to argue for an increase to the amount of causally efficacious content of which we are P-conscious but not A-conscious, which is central to scaffolding the A-P distinction and overflow.
Pain reports under general anesthesia

Block (1995, p. 244) suggests that retrospective reports of P-conscious events under general anesthesia, from patients who had undergone surgery are possible examples of P-consciousness without A-consciousness. Assuming the general anesthetic wipes out A-consciousness but may leave P-consciousness or parts thereof unscathed, Block suggests subjects’ reports of memories of P-conscious experiences (e.g. pains) experienced while they were subjected to a general anesthetic could be veridical.

Aerodontalgia

In his 1995, Ned Block suggests aerodontalgia as an example of P-consciousness without A-consciousness. Aerodontalgia consists in dental pain related to air pressure and was originally discovered in pilots during the second world war. At its core, aerodontalgia relates to possible memory traces of dental work carried out under general anesthesia. Following dental surgery under general anesthesia, stimulations in areas (such as the nasal mucosa or the sinuses) that did not undergo surgery may trigger the experience of pain in the areas that did undergo surgery. That the phenomenon is not linked to the actual current state of the areas at the time of the experience was evidenced by the fact that patients who underwent similar surgeries with the addition of local anesthesia did not report the subsequent pain experiences. Therefore, memory traces seemingly are an apt explanation. Since, ex hypothesi, A-consciousness is knocked offline during general anesthesia, the relevant memory traces presumably are related to P-consciousness.

Hydranencephaly

Sean Smith (2019) has argued for the distinction between A-consciousness and P-consciousness deploying hydranencephaly in infants to illustrate that the two may dissociate in special cases. Hydranencephalic patients are born without a fully developed cortex, which leaves them unable to move (due to lack of motor cortex) and with short life expectancy. Smith argues that hydranencephalics are phenomenally conscious but lack cognitive access. The posit that hydranencephalics are phenomenally consciousness is based on their differentiated affective responses in relation to their primary caregiver versus strangers. The posit that they lack cognitive access, is based on the fact that they are missing (parts of) the cortical areas (such as the prefrontal cortex) thought to underlie A-consciousness. In lieu of ‘normal’ cognitive access, Smith suggests they may have another kind of affective-grounded access, such as the ability to discriminate between the primary caregiver and strangers. Smith thus argues for the A-P distinction based on cases of bodily affect, as opposed to the more conventional approaches, such as visual paradigms.

Molyneux problem and sensory-motor affordances

The Molyneux problem asks what it would be like for a congenially blind person to regain vision as an adult. Hilla Jacobson (2015) considers evidence in which this has been accomplished through surgery. Despite her main aim not being a defense the A-P distinction or overflow, Jacobson delivers an interpretation of the evidence in light of the A-P distinction. The evidence consists in subjects (post-surgery) reporting vivid visual impressions, yet initially having difficulties with identifying objects based on visual information as well as trouble with the localization of objects in space. Supposedly, this shows that some information in the visual phenomenology is not accessible or usable to A-consciousness. What Jacobson defends is essentially the A-P distinction and weak version of overflow that espouses “a notion of a constitutive link between the subject and her phenomenal states that is weaker than her having full-fledged access to those states.” (Jacobson, 2015, p. 1023)

7. A-P distinction: Dissociation at the neural level

Recurrent processing

Recurrent processing is the central element in the most discussed proposal for the neural correlates of P-consciousness, A-consciousness, and their relations, for instance in overflow. The proposal is usually named ‘Local Recurrency Theory’ (LRT). The main evidence supporting the LRT is a series of findings by Victor Lamme (2003; 2004; 2005; 2006; 2010) that is taken to suggest recurrent feedback in the early visual system is necessary for P-consciousness. Lamme found that roughly 100 ms after the onset of a visual stimulus, higher visual areas engage in a feedback loop with the early parts of the visual cortex. According to the LRT the recurrent feedback is the neural correlate of (visual) P-consciousness. The neural correlates of A-consciousness are hypothesized to consist in later processes located in frontal and parietal areas. Recurrent processing has been investigated using a range of different paradigms and techniques (see e.g., Van Gaal and Lamme, 2012), and is often interpreted explicitly in terms of the A-P distinction (both by the original researchers Lamme op. cit., and by philosophers e.g. Block, 2005; 2007). Discussions are ongoing about whether recurrent processing is sufficient for P-consciousness or if it is merely necessary (Sligte, Scholte & Lamme, 2008).

Change blindness

The change blindness phenomenon has been discussed in relation to almost every theory of consciousness and is often mentioned in the literature as an illustrative example of the A-P distinction (Block, 2007; 2011; 2014a; 2015; Jacobson & Putnam, 2016; Lamme, 2003; 2004; Pinto et al, 2017; Sligte et al., 2008; Sligte, Scholte, & Lamme, 2009; Vandenbroucke et al., 2014; Vandenbroucke, Sligte,
Akinetopsia derives from bilateral lesions to V5 in the visual cortex. Ned Block (2005; 2007) deploys this phenomenon as evidence that different phenomenal contents. Like with the other examples of motion experiences in this compilation, Block (2005, p. 46-47) linked to V4 (Sligte, Scholte, and Lamme, 2009). The division of visual short-term memory has been taken to suggest (e.g. Pinto et al., 2005, p. 47). In relation to Akinetopsia, V5 is suggested to be the explicitly evaluates the waterfall illusion as part of an argument that there are distinct neural correlates for P-consciousness and A-consciousness respectively (the A-P distinction). Block suggests that activations in MT/V5 probably further require the involvement of recurrent processing to yield experience as of motion, while activations in the visual system persist. Thus, on the assumption that P-consciousness is distinct from A-consciousness and further that A-consciousness depends on frontal areas, whereas P-consciousness is thought to depend on early sensory regions (as per the local recurrency hypothesis), Block argues that perceptual representations can be consciously experienced without A-consciousness in binocular rivalry.

**Waterfall illusion (motion after effect)**

The Waterfall illusion (or motion after effect) is part of a collection of evidence supporting the hypothesis that the neural correlates of “experiences of as of motion” (Block, 2005, p. 46) reside in the visual cortex, and more specifically in the area of MT/V5. The waterfall illusion consists in a moving afterimage resulting from adaptation in the visual system occurring when looking to a stationary pattern after having looked for a while at a moving pattern. The reason to think the neural underpinnings of this effect, and hence the neural underpinnings of experience as of motion, reside in MT/V5 is that TMS to this area disrupts the moving afterimages (Block, 2005, p. 46). Block suggests that activations in MT/V5 probably further require the involvement of recurrent processing to yield experience as of motion. Thus, the waterfall illusion serves as part of the larger argument for the recurrent processing theory, suggesting that at least visual) consciousness depends on (recurrent) activations in early sensory areas (this is particularly explicit in his 2007). In two papers (Block, 2005, p. 47, and 2007, p. 496), this is laid out as part of an argument to suggest that there are different phenomenal NCCs for different phenomenal contents. Like with the other examples of motion experiences in this compilation, Block (2005, p. 46-47) explicitly evaluates the waterfall illusion as part of an argument that there are distinct neural correlates for P-consciousness and A-consciousness respectively (the A-P distinction).

**Akinetopsia**

Akinetopsia is the inability to perceive motion. That is, subjects with akinetopsia, do not have visual experiences of motion. Akinetopsia derives from bilateral lesions to V5 in the visual cortex. Ned Block (2005; 2007) deploys this phenomenon as evidence that there are specific perceptual NCCs for different phenomenal contents (2005, p. 47). In relation to Akinetopsia, V5 is suggested to be the NCC for motion experiences. Importantly, Block frames his exegesis of the NCC for motion experiences in light of the A-P distinction, arguing that this distinction “helps to make sense of recent results in cognitive neuroscience; we see a glimmer of an empirical case for thinking that they correspond to different NCCs.” (2005, p. 46). Akinetopsia figures as part of a larger argument — based on a collection of empirical phenomena — aiming to show that there are specific phenomenal NCCs (as opposed to access NCCs, see e.g.,
Block 2005, p.47) and that the core neural basis for visual phenomenology is in the back of the head in areas associated with perception (i.e., the visual system, and immediately surrounding areas), as opposed to in frontal areas normally associated with cognition.

**TMS induced phosphenes**

TMS, when applied to the different areas of the visual system, elicits phosphenic experiences. Phosphenes are experienced as brief flashes of light or color. Block (2005; 2007) leverages this in support of the claim that there are different neural bases for different phenomenal contents (Block, 2005, p. 47, and 2007, p. 496). Additionally, when TMS is applied to the area of MT/V5, thought to underpin the experience as of motion, the subjects experience moving phosphenes. Block considers this explicitly in light of the A-P distinction (Block, 2005, p.46–47). Furthermore, the timing of a TMS pulse to MT/V5 was found to influence whether phosphenes were experienced as stationary or moving. This time factor was consonant with the suggested timeframes of recurrent processing in the visual system which further (but indirectly) supports the A-P distinction by virtue of meshing with the preferred hypothesis about the NCCs underpinning A-consciousness and P-consciousness, i.e., the local recurrency hypothesis (Block, 2007, p.496).

**Supraliminal priming during reading**

An experiment by Kouider et al. (2007) has been given an alternative interpretation in light of overflow (Block, 2007, p. 497-8). In the key condition of the experiment, subjects were shown a sequence of visual stimuli, consisting first of a forward mask, then a blank screen, followed by a lower-case word, followed by yet another blank screen, followed by an upper-case word (see Kouider et al., 2007, p. 2021 for details). Subjects were instructed to only look for an upper-case word and ignore any lower-case words. Subjects reported they were aware of the lower-case words, but had difficulty reporting them because they were engaged with the primary task of detecting the upper-case word. According to the interpretation offered in light of overflow, subjects did (P-)consciously experience the lower-case words, but due to attention being allocated to the main task, failed to become A-conscious of the lower-case words, which explains their struggle to report them. In the experiment, fMRI measurements found a difference between an initial subliminal condition, in which both lower- and upper-case words were masked and a supraliminal condition in which there were not. This difference was in the back of the head (i.e., the visual system), which is consistent with the local recurrency hypothesis endorsed by proponents of overflow (Block, 2007, p. 498). Furthermore, the main difference between supraliminal conditions between lower- and upper-case words, when attention was allocated to upper case words, was found to be mainly in frontal coalitions. This seemingly supports the hypothesis Block espouses with respect to the neural correlates of consciousness. Namely that A-consciousness (and reportability) depends on frontal activations.

**Implied motion**

Ned Block (2005; 2007) invokes data (Kourtzi & Kanwisher, 2000) on implied motion in photographs. The data shows V5 is activated when subjects view still photographs if the depicted situation implies motion. Implied motion is part of the data Block deploys to argue that the neural correlates for motion experiences are in V5 (e.g. 2007, p. 496). This, in turn, relates to the more overarching debate about whether the NCCs are to be found in the front or the back of the head. In his interpretation, Block concedes that motion experiences likely require more than just activations in V5 and suggests recurrent feedback as a possibility for one such requirement. Like with Akinetopsia (see above) the implied motion interpretation is cached in light of the A-P distinction (Block, 2005, p.46). Furthermore, the timing of a TMS pulse to MT/V5 was found to influence whether phosphenes were experienced as stationary or moving. This time factor was consonant with the suggested timeframes of recurrent processing in the visual system which further (but indirectly) supports the A-P distinction by virtue of meshing with the preferred hypothesis about the NCCs underpinning A-consciousness and P-consciousness, i.e., the local recurrency hypothesis (Block, 2007, p.496).

8. Overflow

**Partial report superiority (Sperling)**

Colloquially known as the Sperling paradigm, named after the psychologist George Sperling, partial report superiority is among the most debated empirical phenomena in relation to the A-P distinction and overflow. Interpretations exist in both the standard views of the distinction and overflow (Block, 1995; 2007; 2011; 2012b; 2015, Vandenbroucke et al.; 2014), as well as weaker versions (Cheng, 2017). In the paradigm, subjects are presented briefly with a matrix of letters and subsequently asked to report which letters they saw. Under normal conditions, subjects say it felt like they saw all the letters, yet are able only to report 3–5 specific letters, consistent with the consensus capacity of working memory. Sperling’s finding was that if, after the matrix has disappeared, subjects were cued (using a tone) to a specific row in the matrix from which to report the letters, they retained the ability to report between 3 and 5. That subjects retain this ability — despite not knowing which row they would be required to report — indicates that information about the identity of all the letters is available prior to the report. The A-P distinction and overflow explanation of this takes the subjects’ reports as veridical, i.e. they experienced seeing (were P-conscious of) all the letters in the matrix. However, because A-consciousness ex hypothesis involves working memory and has limited capacity, not everything in P-consciousness can be A-consciousness at a given time (overflow). P-consciousness of the letters at times has been associated with (visual) iconic memory. Mack et al. (2015) suggested that a variation of the paradigm showed that iconic memory requires attention. In reply to this, Bachmann and Aru (2015, p. 73) offer an alternative interpretation of the Mack et al. variation, which suggests that “the attentional manipulation only influenced the access to iconic memory […]”, attention had its effect not on the existence of phenomenal experience of iconic memory but on the readout from
iconic memory.”

Inattentional blindness

Inattentional blindness is a phenomenon in which subjects fail to notice unexpected stimuli, if their attention is focused on another task. When asked whether they noticed the unexpected stimuli, subjects often cannot report anything about them. In his (2004), Victor Lamme interprets this by deploying the A-P distinction. He posits the lack of attention as the key to understanding the subjects’ failure to notice the unexpected stimuli. Lamme suggests that subjects nevertheless are P-conscious of the unexpected stimuli, but that the experience is fleeting when unattended. A-consciousness — and with it, reportability — is facilitated by attentional processes, supposedly through the storage of items in working memory (Lamme, 2004, p. 864). In another interpretation of inattentional blindness, Usher and colleagues (2018) suggest that it is due to the fact that the unexpected stimuli are irrelevant to the assigned task, that subjects fail to notice them. Usher et al. offer an interpretation of inattentional blindness in terms of parsing incoming stimuli for task relevance, which they propose serve as a gateway to conceptual categorization i.e., the activation of semantic representations (Usher et al., 2018, p. 8), which in turn is needed for A-consciousness and verbal report. According to this interpretation, P-consciousness overflows reportability because irrelevant stimuli are not categorized, preventing them from becoming A-conscious. If this is the case, then phenomenal overflow would — at least in part — derive from a categorization bottleneck.

Color diversity

The color diversity paradigm—made famous by Marius Usher and Zohar Bronfman (Usher et al. 2018; Bronfman et al., 2014)—is a variant of the Sperling paradigm. The core idea is that the letters in the Sperling grid are different colors, and the subjects are subsequently tasked with judging whether the diversity of colors was ‘high’ or ‘low’. Subjects made the color diversity judgment, either in the row cued for the main task, or in the set of uncued rows. Despite attention being allocated to the Sperling task, subjects did well on the color diversity judgments even for the uncued rows. Their interpretation is cast as a mild version of overflow suggesting that “phenomenal consciousness overflows encoding into WM [Working Memory] and report” (Usher et al., 2018, p.9). The reason given in support of the assumption that the colored array was P-conscious is “that the detection of this nonattended information is based on a conscious experience of the underlying elements, as observers were unable to (subliminally) guess the color diversity of the array when they reported having no experience of the colors” (Bronfman et al. 2014, p. 1401). Block (2014d) applied the color diversity finding in light of whether visual consciousness is rich or sparse (a debate between the proponents of the A-P distinction and overflow, and a conglomerate of opponents endorsing other theories). Framing the interpretation of the color diversity data in contrast to workspace theories, Block (2014d, p.446) concludes that the data ‘reveals that there must have been conscious awareness of specific colors beyond the limits of the global workspace because a trace of that conscious awareness in the form of a diversity judgment can enter the global workspace for free.” A similar interpretation is found in Jacobson (2015) who says (p.1032–1033): “[…]the number of experiences of individual colors to which the subject has access is subject to the familiar limitations of working memory — about three or four items; but the relevant judgments of color-diversity can be based on representations – and moreover, on phenomenal representations – of many more colors” (Italics from original).

Amodal completion

The amodal completion paradigm that Usher et al. (2018, p. 9) leverage to argue that categorization or knowledge activation is more sensitive to task relevance than P-consciousness, uses incomplete Kanizsa-like inducers (that do not combine to form an illusory shape) surrounding a non-illusory rectangle. The subjects are tasked with judging whether the rectangle is vertical or horizontal. On the critical manipulation, the Kanizsa inducers actually combine to form an illusory rectangle and the non-illusory rectangle is removed. The finding is that the subjects still perform well when judging whether the Kanizsa rectangle was vertical or horizontal. This is interesting to Usher et al. when combined with evidence that amodal completion (i.e., the illusory Kanizsa rectangle) only seems to work when the inducers are conscious. This suggests that the subjects were conscious of the inducers, despite these being task irrelevant up until that point. Combining this with their interpretation of Irrelevance induced blindness (see below), Usher et al. conclude that subjects are P-conscious of the inducers (which supposedly generalizes to task irrelevant stimuli in general), because (on the assumption that amodal completion only works when the inducers are conscious) this follows from the subjects’ performance in judging the orientation of the illusory rectangle. The upshot, and the tacit aim of Usher and colleagues, is that when considering other empirical data, in which subjects show poor ability to report on task irrelevant stimuli, we cannot conclude that they were not P-conscious of the stimuli.

Divided attention paradigms

In divided attention paradigms, subjects perform an attention demanding main task, but the measure is performance on a secondary visual discrimination task. Seemingly, subjects can perform several types of visual discrimination despite having their attention allocated elsewhere. The argument assumes that attention is associated with A-consciousness and has a limited capacity (presumed to be occupied by the attention demanding primary task) and suggests that if subjects can perform the secondary visual discrimination task, this ability cannot be explained by the deployment of A-consciousness. This, supposedly, leaves the explanation that the performance is driven by perceptual experience. Usher et al. (2018, p. 2) says the argument is “aimed to show that conscious experience
can take place outside the focus of attention.” The aim of Usher et al., (2018, p.9) in the pursuit of which the divided attention data is deployed, is to argue in favor of overflow (and against a no-overflow view).

**Attentional blink**

The attentional blink phenomenon concerns subjects’ ability to detect a target item in a series of visual stimuli in cases where it is preceded by another target item at specific time intervals. Presumed to drive this phenomenon is attention being occupied with the first target, and therefore missing the second target. Since the ‘blink’ only occurs during a limited time span, varying the interval between the targets manipulates the likelihood that subjects will detect the second target. Block (2007, p. 497) suggests that activations in the back of the head (i.e., the visual system) related to missed targets often may be almost as strong as those of the reported targets, even though the former fail to activate frontal areas (or workspace ignition) associated with attention and report. In the paradigm Block considers, subjects were asked to indicate how visible the second target was, and their responses tended toward either maximally visible or completely invisible, which suggests an all-or-nothing process. On Block’s interpretation, subjects’ failure to detect the second target does not mean they did not consciously see it. On the assumption that phenomenology is driven by early sensory areas, and strength of activation related to the missed target rival those of the first target, subjects would be consciously seeing the second target, but must have been prevented from reporting it for some reason, viz. because it was not accessed. Supposedly, this is related to the limited capacity of A-consciousness, and the fact that it was already engaged with the first target. The conclusion (p. 498) is that this supports overflow.

**Irrelevance induced blindness**

Arguing against the position that P-consciousness merely overflows reportability (as opposed to A-consciousness), Usher et al. (2018, p.8) deploy irrelevance induced blindness. In irrelevance induced blindness subjects perform a relatively trivial task (e.g. judging which circle is largest). During the task, an irrelevant property (e.g. color) of the task relevant stimulus is changed. Usher et al., propose that subjects perform poorly when asked to subsequently report the color. Usher et al., suggest that the subjects perform poorly “not because they are phenomenally unconscious of a simple stimulus which is fully attended, but rather because, being task-irrelevant, the stimuli are not categorized, or if they are categorized, they are then rapidly forgotten.” (Usher et al., 2018, p. 9).

**Identity crowding**

In two papers (Block, 2012a; 2014b), Block deploys identity-crowding to argue that “there can be conscious perception of an object without attention to that object” (Block, 2012a, p. 170). The crowding phenomenon derives from visual stimuli with multiple items, where the “critical spacing” for crowding is defined as the spacing between items allowing for 80% identification (Block, 2014b, p. 162). Then, Block delineates the grain of attention by the angle from the eye at which the subjects no longer can determine whether an object in a crowded visual stimulus changed (2012a, p. 174). Block argues that identity-crowded items can be consciously detected, identified, differentiated, and discriminated even while they are below the grain of attention. Supposedly, because the identity-crowded items are below the grain of attention, these capacities of subjects cannot rely on attentional processes (i.e., A-consciousness), but instead must depend on (at least partly) perceptual processes (2014b, p. 162). Thus, perceptual judgment (A-consciousness) has limitations, when it comes to identity-crowding, and perceptual experiences (P-consciousness) is suggested to underpin the ability to detect, differentiate, identify, and discriminate crowded items.

**Aphantasics’ performance on retro-cue paradigms**

In two papers, D’Aloisio-Montilla (2017; 2018) argues that the ability of aphantasics to perform well on change detection paradigms counts in favor of overflow. Aphantasia consists in a condition where individuals lack the ability to generate voluntary mental imagery. The basic premise for D’Aloisio-Montilla is that no-overflow accounts must appeal to an internal image to explain the results of retro-cue paradigms (Landman et al., 2003). From this, D’Aloisio-Montilla argues that the fact that the performance of aphantasics is on par with that of neurotypical subjects means appealing to internal imagery to explain performance in the retro-cue paradigms is not warranted. Since appealing to internal imagery cannot explain the performance of aphantasics, D’Aloisio-Montilla argues it is unlikely that internal imagery is driving the performance in neurotypical subjects. If we cannot appeal to internal imagery, D’Aloisio-Montilla argues, this supports the competing interpretation, viz. overflow accounts that propose that performance in retro-cue tasks is driven by the retro-cue allowing delayed access to the phenomenal experience of the original stimulus presentation.

**Metacognition for unattended visual representations**

Vandenbroucke and colleagues (2014) investigated whether the introspective feeling of seeing more than can be attended to is veridical or illusory (the so called rich-sparse debate, which is a proxy for the debate on overflow). They (Vandenbroucke et al., 2011) compared subjective confidence ratings for different stages of visual memory while keeping objective performance stable. They found that the metacognitive accuracy, understood as subjects’ ability to correctly judge whether or not their visual judgments were accurate on a change detection task, was similar or higher for sensory memory than for working memory. This supposedly indicates that there
are more items in sensory memory than in working memory, which maps reasonably onto overflow. In line with other interpretations (e.g., Pinto et al., 2017a,b) associating sensory memory with P-consciousness and working memory with A-consciousness, Vandenbroucke and her team conclude that "unattended, sensory memory items are a meaningful part of visual experience". (Vandenbroucke et al., 2014, p. 870).

**Chromatic perception**

John Beeckmans (2009) has proposed chromatic color vision as evidence supporting overflow. Beeckmans suggests that A-consciousness primarily operates using mereologically superordinate concepts that are derived from the rich and complex chromatic information available in the early visual system. Because the mereologically superordinate concepts deployed in A-consciousness, ex hypothesi, are significantly more coarse-grained than the full chromatic spectrum present in the visual system, Beeckmans proposes (2009, p. 925) that "chromatic content of visual phenomenality is in general far richer than is its accompanying conceptualization in A-consciousness". Beeckmans (2009, p. 923) supports this conclusion by reference to a Sperling type experiment using a large set of color discs, in which subjects subsequently had to report on their visual experiences. In a nutshell, Beeckmans’ argument is that because our color concepts of highest specificity (e.g. "dark indigo") encompass a multitude of finer grained shades, which are reflected in our visual system (and hence P-consciousness), if A-consciousness depends on these concepts, then necessarily there are P-conscious contents cannot be accurately reflected in A-consciousness.

**Random dot stereograms**

To argue that there is phenomenal persistence in at least some parts of vision, Ned Block (2007, p. 490) invokes Engel’s (1970) use of random dot stereograms. In the experiment, subjects viewed grids of small black or white squares through stereograms. Upon systematically shifting sets of the squares to one side, in the presentation to one eye, subjects perceived apparent depth in the scene. By varying the timing of the shifts in the stimuli, it was shown that maintaining the perception of depth was possible with as much as 300 ms temporal asynchrony in the stimulus presentations. This suggests that phenomenal persistence (i.e., the experience of depth), cannot be explained exhaustively in terms of low-level visual properties, because it persists significantly longer than these. The argument for phenomenal persistence is part of the larger argument for overflow, since it serves to scaffold the interpretations of other empirical data, most notably the partial report paradigms. This is because what is posited to explain e.g., partial report superiority, is exactly the persistence (or memory trace) of phenomenally conscious states.

**Olfactory perception**

Stevenson and Mahmut (2014) asked subjects to judge features of odors under different conditions. Two variables were deployed. The first variable was whether the odor was present or absent, i.e., whether the subjects in their judgments could rely on perception or had to rely on recall. The other variable was the features of the odor. In “congruent” trials subjects judged the same feature during perception, but also in olfaction, and cast (2014, p.210) their findings explicitly in light of the A-P distinction.

**Differences between iconic memory and visual working memory**

Quilty-Dunn (2020) discusses differences in the format and features between iconic memory and visual working memory. Iconic memory exhibits holistic feature binding, whereas visual working memory does not. The conclusion he aims at, is that iconic memory stores unconceptualized visual icons, which contrasts with what he calls discursive representations in visual working memory. Moreover, he posits that visual icons are informationally richer than discursive representations in visual working memory. He concludes that these differences provide “a solid foundation for arguments […] for the thesis that phenomenal consciousness overflows cognitive access”. (Quilty-Dunn, 2020, p.18).

**9. Examples of validation**

In this section, we turn to the importance of the validation step in IBE. As mentioned, the aim of the validation step is to check, on a case-by-case basis, if the suggested interpretation of an empirical finding can provide the alleged support for the view considered.

For a positive example, consider the relation between partial report superiority and the overflow view. Note that although perhaps this is the most discussed finding that has ever been claimed to be part of the empirical basis of the overflow view the debate is not over whether the original interpretation of this phenomenon is correct or whether it is compatible with the overflow view itself, but rather whether there can be alternative explanations, i.e. whether a no-overflow theorist could accommodate this finding into his/her framework. It is a consensus in the literature that what the Sperling paradigm shows is that subjects presented a matrix of letters can report 3–5 letters from any row if that row is cued shortly after the disappearance of the original stimulus. And the picture offered by
the overflow view, namely that subjects are P-conscious of all the letters in the matrix (whereas they are A-conscious of only a fraction of them) is certainly compatible with this finding.

However, the validation step is important exactly because neither the correctness of the original interpretation nor the compatibility of the finding in question with the theoretical claim of the view considered is a trivial matter. To see why, consider the claim that aphantasics perform at the same level in so-called retro-cue paradigms as individuals with strong imagery ability (D’Aloisio-Montilla, 2017; 2018). Recall that aphantasia is a condition where individuals are unable to generate voluntary mental imagery. Relying on this fact, D’Aloisio-Montilla runs the following argument. First (premise 1), he claims that any sound account of the retro-cue task must satisfy that subjects consciously experience a successfully reported, cued item. Second (premise 2), he argues that no-overflow theorists satisfy this claim by appealing to a type of internal imagery. Third (premise 3), he introduces evidence (Keogh & Pearson, 2014) that aphantasics lack the ability to generate this imagery. Fourth (premise 4), he cites findings by Keogh and Pearson (2011) that he interprets as showing that aphantasics perform equally well in retro-cue trials. Finally, since on no-overflow accounts good performance in retro-cue tasks requires mental imagery (premises 1 & 2) that aphantasics lack (premise 3), D’Aloisio-Montilla concludes that the good performance of aphantasics in retro-cue tasks (premise 4) cannot be accounted for within the framework of the no-overflow view.

Of course, for all this to be able to provide support for the overflow view, it needs to be the case that the overflow theorists have no similar problem with accounting for this phenomenon. According to D’Aloisio-Montilla since overflow theorists appeal to a conscious experience of the whole array which can satisfy premise 1 without relying on mental imagery, premise 3 poses no problem for overflow theorists in their account for premise 4.

Thus — it seems — the crucial premise in this argument is premise 4: D’Aloisio-Montilla’s interpretation that the Keogh and Pearson (2011) findings show that aphantasics perform on par with individuals with normal imagery abilities in the retro-cue paradigm. Evaluating this interpretation is key to validating whether the aphantasics performance on retro-cue paradigms can count as empirical support for the A-P distinction and overflow. Closer reflection reveals two problems with the interpretation that aphantasics can perform well in the retro-cue paradigm: first, the subjects in the original experiments were not diagnosed with aphantasia, and second, the experimental paradigm used was not a classical retro-cue paradigm. Let us explore these two problems in order.

The original Keogh and Pearson (2011) study investigated individuals with varying imagery ability. They had to hold fixation while seven Gabor patches were presented in the periphery around the fixation point. After a retention period with no stimulus, a line cue pointing to the location of a previous Gabor patch (randomized) was presented. Participants needed to indicate whether the orientation of a subsequently presented Gabor patch in the cued position was rotated clockwise or counterclockwise compared to the original Gabor patch. According to the relevant finding, individuals with poor imagery ability performed similarly to those with strong imagery ability. However, no indication whatsoever is given regarding whether the individuals in the poor imagery ability group lacked the ability to generate voluntary mental imagery altogether (i.e., if they were aphantasics). Noting this shortcoming of his original argument, in a subsequent paper D’Aloisio-Montilla (2018) argued that no reference to aphantasia was required for his argument to go through. If individuals with poor imagery ability perform equally well in retro-cue trials then imagery ability does not correlate with the performance in these trials, which is what the no-overflow theorist would predict (given premise 2).

Although shifting focus from aphantasia to the lack of correlation between imagery ability and performance seems to be the right move to deal with the first problem effectively and can provide a more faithful interpretation of the actual experiment, it cannot circumvent the second problem. The second problem is that the change detection paradigm used by Keogh and Pearson (2011) was crucially different from the classical retro-cue paradigms deployed in the overflow debate, where participants need to detect changes in a complex stimulus at a position indicated by a cue that is presented not less than 1 sec after the stimulus offset. In the Keogh and Pearson (2011) experiment, however, the cue occurs significantly sooner than 1 s—following only after 20 ms, 40 ms or 400 ms retention periods. This is why the original authors themselves refer to it as an iconic memory task. It is a crucial difference, as for an iconic memory paradigm D’Aloisio-Montilla’s original argument does not go through. Premise 2, i.e., the requirement that no-overflow theorists need to rely on mental imagery to account for the performance of the subjects applies only to cases that cannot be explained by reference to iconic memory. In iconic memory paradigms (as D’Aloisio-Montilla (2017) himself acknowledges) utilizing mental imagery is not necessary for a strong performance. When the retention period is shorter than 500 ms (Coltheart, 1980; Graziano & Sigman, 2008) cues can help move content from an unconscious iconic memory store to WM (see classical no-overflow interpretations of the Sperling experiment; Phillips, 2011; Stazicker, 2011).

That is, D’Aloisio-Montilla’s original argument, even if valid, is unsound: premise 4 is not true. More precisely, the kind of experimental paradigm that is needed for the truth of premise 2 of the argument (retro-cue paradigm) is different from the experimental paradigm that premise 4 refers to (iconic memory paradigm). Having said all this, note that the conclusion above regarding the aphantasia-based argument against the no-overflow view is relevant only if one wants to evaluate if the Keogh and Pearson (2011) finding can be part of the empirical basis of the no-overflow account. From our current perspective, the question is if the suggested interpretation of this experimental finding is unproblematic and whether it is compatible with the overflow view—i.e., if the empirical basis of the overflow position contains the phenomenon that D’Aloisio-Montilla describes. The answer to this question (and thus the conclusion of the validation step) is negative: “equal performance of aphantasics in retro-cue tasks” is not something that one should think of as a phenomenon that supports the overflow view, simply because there is no such phenomenon. What the original Keogh and Pearson (2011) study demonstrates is the existence of a different phenomenon, namely that there is no difference in the performance of individuals with poor and strong imagery ability in iconic memory tasks. While the phenomenon so characterized can be part of the

---

[8] Imagery ability was assessed by a binocular rivalry paradigm based on the observation that when one of two rivalry patterns was imagined that pattern had a higher probability of being dominant during a subsequent brief rivalry presentation (see Pearson et al., 2008).
empirical basis of the overflow view, the phenomenon as characterized by D’Aloisio-Montilla’s description is not.\(^9\)

10. Concluding remarks

The ability of empirical evidence to move forward the science of consciousness requires a clear view of how it is supposed to do so. We have posited that the process of inferring to the best explanation is likely to be necessary at some point. This posit rests on the assumption that a process of eliminating theories through falsification will not suffice to settle which theory of consciousness accurately reflects the way the world is. Among the reasons to think falsification will fail to ultimately settle this issue are arguments that some theories seem unfalsifiable, and that (at least) our current technology and methodologies are insufficient (see introduction). Another — more principled — reason is that many of the concepts deployed in theorizing about consciousness do not map straightforwardly onto the kind of data generated by the brain sciences, and vice versa (Overgaard & Kirkeby-Hinrup, 2021). To boot, much work in the field is tacitly framed as part of this IBE process, and sometimes explicitly stated to be.

We have proposed a rational reconstruction of what the IBE process — to our minds — should look like, when applied to empirically minded consciousness studies. The majority of work in the field (for uncontroversial reasons) has been concerned with the first (assimilation) of the four steps involved in the IBE process. We have here highlighted the importance of the three additional steps and demonstrated that step two (compilation) and step three (validation), while cumbersome, are feasible. We did this by delivering a compilation of empirical evidence deployed in support of the two central claims of Ned Block, namely the A-P distinction and the overflow hypothesis, and considering the evidence proposed by D’Aloisio-Montilla suggesting that aphantasics performance on the retro-cue task supported the overflow view. Since meeting the preconditions to even start working on step four (comparison) are still some way into the future, we cannot give an example of this step here.

Finally, whether or not one endorses our rational reconstruction of the IBE process, we are confident that the compilation provided in this text and an eventual online open-access database will be a valuable resource to assess the debates between competing theories of consciousness. If one wants to take seriously the potential for empirical evidence to move forward debates in consciousness studies, there is a need to gather, organize, validate, and compare evidence. This is true whether one considers it as part of a larger IBE process, or not. Neglecting this kind of work is detrimental to the overarching project of understanding consciousness through application of empirical data, as well as the particular debates in which the empirical data is applied. To move forward the science of consciousness through the application of empirical evidence, we should want to ensure that the specific instances of application, and the overall methodology of application stand up to scrutiny.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

AKH is funded by the Swedish Research Council (grant# 2018-06595). PF has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754513 and The Aarhus University Research Foundation.

References

Armstrong, D.M. (1968). A materialist theory of the mind: Reprinted Routledge 2002.
Bach-y-Rita, P., & Collins, C. C. (1971). Sensory substitution systems using the skin for the input to the brain. Journal of the Audio Engineering Society, 19(5), 427–429.
Bachmann, T., & Aru, J. (2015). Comments on how Mack et al. (2015) (do not) see iconic memory. Consciousness and Cognition, 34, 73–74.
Beeckmann, J. (2009). How chromatic phenomenality largely overflow its cognitive accessibility. Consciousness and Cognition, 18(4), 917–928.
Block, N. (1995). On a Confusion About a Function of Consciousness. Behavioral and Brain Sciences, 18(2), 227–247.
Block, N. (1996). Mental paint and mental latex. Philosophical Issues, 7, 19–49.
Block, N. (1998). How to Find the Neural Correlate of Consciousness. Royal Institute of Philosophy Supplement, 43, 23–34.
Block, N. (1999). Sexism, Racism, Ageism, and the Nature of Consciousness. Philosophical Topics, 27, 1–20.
Block, N. (2005). Two neural correlates of consciousness. Trends in Cognitive Sciences, 9(2), 46–52.
Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. Behavioral and brain sciences, 30(5-6), 481-499; discussion 499-548.
Block, N. (2008). Consciousness and cognitive access. In Proceedings of the Aristotelian Society (Vol. 108, No. 1 pt.3, 289–317.
Block, N. (2009). Comparing the major theories of consciousness. In M. S. Gazzaniga, E. Bizzi, L. M. Chalupa, S. T. Grafton, T. F. Heatherton, C. Koch, J. E. LeDoux, S. J. Luck, G. R. Mangan, J. A. Movshon, H. Neville, E. A. Phelps, P. Rakic, D. L. Schacter, M. Sur, & B. A. Wandell (Eds.), The cognitive neurosciences (pp. 1111–1122). Massachusetts Institute of Technology.
Block, N. (2011). Perceptual consciousness overflows cognitive access. Trends in Cognitive Sciences, 15(12), 567–575.
Block, N. (2012a). The grain of vision and the grain of attention. Thought: A Journal of Philosophy, 1(3), 170–184.
Block, N. (2012b). Response to Kouider et al.: Which view is better supported by the evidence? Trends in Cognitive Sciences, 16(3), 141.

\(^9\) For a recent argument that can lead to another negative example of validation, consider how Phillips (2021) calls the standard interpretation of blindsight—on the basis of which it is usually claimed to be part of the empirical basis of the overflow view—into doubt.
A. Kirkeby-Hinrup and P. Fazekas

Pinto, Y., Vandenbroucke, A. R., Otten, M., Sligte, I. G., Seth, A. K., & Lamme, V. A. (2017). Conscious visual memory with minimal attention. *Journal of Experimental Psychology: General, 146*(2), 214.

Pitts, M. A., Lutsysyyna, I. A., & Hillyard, S. A. (2018). The relationship between attention and consciousness: An expanded taxonomy and implications for ‘no-report’ paradigms. *Philosophical Transactions of the Royal Society B, 373*, 20170348.

Quilty-Dunn, J. (2020). Is iconic memory iconic? *Philosophy and Phenomenological Research, 101*(3), 660–682.

Rensink, R. A., O’Regan, J. K., & Clark, J. A. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368–373.

Sebastian, M. A. (2014). Not a HOT dream. In R. Brown (Ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience* (Vol. 6, pp. 415–432). Netherlands: Springer.

Shoemaker, S. (1982). The inverted spectrum. *Journal of Philosophy, 79*(7), 357–381.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261–267.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2008). Are there multiple visual short-term memory stores? *PLOS One, 3*(2), Article e1699.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2009). V4 activity predicts the strength of visual short-term memory representations. *Journal of Neuroscience, 29*(23), 7432–7438.

Smith, S. M. (2019). Phenomenal overflow, bodily affect, and some varieties of access. *Review of Philosophy and Psychology, 10*(4), 787–808.

Stevenson, R. J., & Mahmut, M. (2014). Evidence that phenomenal olfactory content exceeds what can later be accessed. *Consciousness and Cognition, 30*, 210–219.

Tallon-Baudry, C., Campana, F., Park, H.-D., & Babo-Rebelo, M. (2018). The neural monitoring of visceral inputs, rather than attention, accounts for first-person perspective in conscious vision. *Cortex, 102*, 139–149.

Usher, M., Bronsman, Z., Talmor, S., Jacobson, H., & Eitam, B. (2018). Consciousness without report: Insights from summary statistics and inattention ‘blindness’. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1755), 20170354.

Van Gaal, S., & Lamme, V. A. (2012). Unconscious high-level information processing: Implication for neurobiological theories of consciousness. *The Neuroscientist, 18*(3), 287–301.

Vandenbroucke, A. R., Sligte, I. G., Barrett, A. B., Seth, A. K., & Lamme, V. A. (2014). Accurate metacognition for visual sensory memory representations. *Psychological Science, 25*(4), 861–873.

Vandenbroucke, A. R., Sligte, I. G., & Lamme, V. A. (2011). Manipulations of attention dissociate fragile visual short-term memory from visual working memory. *Neuropsychologia, 49*(6), 1559–1568.

Yaron, I., Melloni, L., Pitts, M. and Mudrik, L., 2021. The Consciousness Theories Studies (ConTraSt) database: analyzing and comparing empirical studies of consciousness theories. bioRxiv.

Weisberg, J. (2014). Sweet dreams are made of this? A HOT response to Sebastián. In R. Brown (Ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience* (Vol. 6, pp. 433-443): Springer Netherlands.