Rich Perceptual Consciousness and Illusion-based Interpretation

Zhiwei Yang, Peking University

Draft of June 4, 2023

Abstract: Phenomenal overflow has garnered significant attention in the field of consciousness research due to its potential to shed light on the relationship between phenomenal consciousness and access consciousness, it postulates that the phenomenal consciousness of individuals may exceed their differential access consciousness, because of a phenomenon observed in certain scientific psychological experiments where subjects claim to perceive all that is presented to them but are only able to report part of it. This phenomenon has been challenged by critics who argue that the apparent perceptual richness is merely an illusion. However, upon a reexamination of relevant experimental data, it is evident that the illusory interpretation is at least partially unreasonable, as it is facing similar issues to those of overflow, and the debate goes on and on.

Keywords: Phenomenal overflow, Perceptual richness, Illusion, Consciousness

1. Introduction

In order to gain a better understanding of consciousness, the notion of allowing multiple properties to serve as the essence of consciousness has been proposed. This is expressed through the differentiation of consciousness into various types based on the essential properties it involves, primarily phenomenal consciousness which represents the experiential or phenomenal aspect of consciousness, and access consciousness which represents its functional aspect. According to this proposition, different types of consciousness represent distinct essential properties and are independent of each other. However, the reliability of this distinction remains to be verified empirically, with the most

*Correspondence E-mail: allenminesky@gmail.com
influential proposition being the *phenomenal overflow*, which is refers to the fact that the capacity of phenomenal consciousness exceeds that of access consciousness, indicating the existence of a portion of consciousness that cannot be cognitively accessed. The Sperling paradigm is widely cited as evidence for this separation, with subjects presented with a complex stimulus consisting of three sets of four letters, they consistently reported seeing more letters than they could subsequently recall, supporting the argument for phenomenal overflow, which maintains that not all perceptual consciousness can be accessed. However, this interpretation has been challenged by those who argue that subjects have cognitively accessed all of their perceptual consciousness and that the so-called perceptual richness they experience is simply an *illusion*. Known as the illusory interpretation, this would cast doubt on the results of the Sperling paradigm as evidence for phenomenal overflow and call into question the idea that consciousness can be separated into different types. The validity of the overflow interpretation remains a subject of ongoing debate within the academic community.

This paper adopts a neutral stance but deliberately situates itself within the overflow framework to facilitate subsequent analysis. After presenting the proposition of phenomenal overflow and discussing its main challenges, particularly the illusory interpretation, the paper examines whether the illusory interpretation, which aims to explain the data on which overflow is based, faces similar issues to those faced by proponents of overflow. Specifically, whether the data on which their interpretation depends can also be interpreted by overflow. The analysis concludes that similar issues appeared, it can be cogently argued that both interpretations retain some degree of explanatory power for the opposing datasets, and both are plagued by certain internal inconsistencies that continue to fuel the ongoing debate.

### 2. Separation and overflow

The concept of consciousness is ambiguous, it points to too many different things, this idea (Block, 1995) comes from similar situations in the history of science, such as the paradox founded by the Florentine experimenters in the 17th century who used the word *degree of heat* to denote *temperature* and *heat*, for consciousness, it is also possible to face the above situation when different concepts are treated as a single one, a most typical one is that people seemly tend to offer functional descriptions
regarding consciousness processing (Baars, 1993; Crick and Koch, 1995; Kouider and Dehaene, 2010). One of the tasks of philosophy is precisely conceptual clarification, and in order to avoid the confusion caused by the use of concepts, consciousness is distinguished into different kinds according to the properties they involve, among which the representative ones are phenomenal consciousness, which involves the experiential properties, and access consciousness, which involves the functional one. There are different ways to characterize these two, and Ned Block, the originator of the distinctive concept, uses synonyms to do it†. The synonym for phenomenal consciousness is experience, for example, when I am sitting on a couch in a coffee shop typing on my laptop, I feel not only the feedback of my fingers on the keyboard, and the bright colors of the images but also the warmth of the air-conditioning breeze on my face, this sort of experiential content, which can be described by what-its-like-ness‡, is considered to be phenomenal consciousness. On the other hand, the synonym for access consciousness is awareness, which is thought to be used to share the aforementioned content with other cognitive systems in charge of reasoning and rational control of action and reporting (Block, 1995, 227). An example of this is when I accidentally get burned by a fire while using the stove for heating, and I generate experiential content about the pain, and it is this conscious content that causes me to scream and urgently look for a band-aid, I am in a mental state that can be understood as access consciousness.

As Block (1995, 233-234) said, these two consciousnesses can be conceptually separated from each other, the separation was justified immediately by some thought experiments he tried to propose§, and have subsequently received widespread attention in fields such as philosophy and scientific research, although some agree with them and some do not. Empirical evidence seems to be important in this case and then, a proposition called phenomenal overflow emerges, which claims that a portion of phenomenal consciousness as overflowed can be proven to exist independently of access consciousness by giving a rational interpretation for some data from the scientific experiment (Block,

†The main reason for using a synonymous description is that it is difficult to define consciousness in any remotely noncircular way, almost any words used to define it are inevitably related to consciousness itself (Block, 1995, 230).

‡what-its-like-ness also used in the same way as qualia, phenomenal character, subjective character, etc. In addition, it should be noted that Block also takes experiential properties of desires, emotions and so on as P-conscious properties, but this is not highly relevant to the topic of this paper, so I skipped this part (Block, 1995, 230).

§The most typical are the (super and super-duper) blindsight and pneumatic drill cases, which as a reason to believe in access consciousness without phenomenal consciousness and the opposite in the conceptual level, respectively. As the paper examines relevant evidence and arguments on an empirical level, the soundness of these thought experiments is not discussed (Block, 1995, 233-235).
As the most famous one, subjects were shown a three times four array of letters with a tachistoscope for a very short time, immediately afterward, they were asked to report the contents of all the letters they saw (so-called all-report method), it was found that subjects could generally report only three to four one in the array at random, just about fifty percent of the wholly presented. Astonishingly, even with limited reporting ability, subjects still claimed to have a perceptual richness of the whole letters, perhaps resembling what Fodor (2007) calls “a general phenomenal intuition.” Could it be that the subjects actually saw far more letters than they could report, but perhaps the images of the remaining letters had completely faded in the brain by the time they reported the third or fourth letter, causing them to show limited reporting ability? Unlike the previous method, for finding out if they had seen in a brief exposure to a complex visual stimulus, in the new partial-report method, each of the three rows of letters in the array was given a high, medium, and low tone cue, and after the array disappeared, the subjects were instructed to report the letter content of the corresponding row by playing the specific row tone. New results showed that subjects were indeed successful in reporting almost letters, leading Sperling (1960) to conclude that the subjects had seen almost all of the letters during the short time the array was presented and that partial letters could not be reported afterward for some reason.

One reason Sperling speculated is that there might be a short-lived sensory memory (iconic memory), it is a part of the visual memory system that its content is not held for long periods of time and is only transferred to a more limited capacity working memory, which is a capacity-limited cognitive system with a capacity limit similar to the number of subject reports (approximately 3-4) (Block et al., 2014), to support subsequent reports with the engagement of attention (Sperling, 1960, 1). Block has mounted a more general argument that there is more in our conscious phenomenology than what we can access. For subjects, they have phenomenally conscious experience about the perceptual content or representation they were shown, and the capacity of perceptual content-equaled

---

4In this paper, I assume that the report ability is credible, for issues related to the validity of subjective reports, see (Block, 2019; Michel, 2022; Richards, 2015).

5There is some debate about the capacity limits of working memory and its relationship to cognitively accessed content, but at least both sides of the debate about whether the overflow proposition holds tacitly share common assumptions, See (Gross and Flombaum, 2017).
iconic memory” is greater than that of cognitively accessed-equalized working memory, only a portion of perceptual content which entered working memory (with the help of attention††) can be reported, the rest, i.e. the part that has been overflowed, can’t be reported, although they think that it is also phenomenally conscious experience as the same as the reported part (Block, 2007a,b, 2011, 2014a; Burge, 2007; Tye, 2006). But is this the case and can the unreported perceptual content be consciously experienced? Probably not.

3. Two different illusion-based interpretations

Someone says that only the reportable content accessed by subjects is what is experienced consciously with the help of attention, while they cannot experience the unreported part in the state of consciousness (Brown, 2012; Cohen and Dennett, 2011; de Gardelle et al., 2009; Kouider and Dehaene, 2010; Lau and Rosenthal, 2011; Phillips, 2011, 2016; Stazicker, 2011). What supports this view are two kinds of interpretations based on different illusions, the generic(gist-like) illusion, and the fragmentary illusion, both of which try to interpret the perceptual richness as a consequence of an illusion.

According to the generic illusion, unreported letters are represented in a generic, gist-like, or merely determinable manner, e.g., as a series of letter-like forms as opposed to specific letters, without the engagement of attention (Cohen and Dennett, 2011; Phillips, 2011, 2016; Stazicker, 2011). “Rather than seeing the cue as deciding which items from rich conscious experience are stored in working memory, we can think of it as deciding which unconscious representations rise to consciousness (Cohen and Dennett, 2011, 360).” Subjects may have (visual) perceptual representation (content) of uncued rows but cannot experience specific one consciously, until with the help of the tone cue, providing top-down attention, helps them generate consciousness about the letters cued, which is used for reporting. Before the cue, the letters can be experienced consciously in a gist-like manner instead of a specific manner. Perhaps the subjects simply experienced a “refrigerator-light illusion”, which is based on a mistake that a child might make: a child would think that the refrigerator light is

---

*The main reason for using the two terms interchangeably is that the brain region where iconic memory resides is a candidate for where the neural basis of conscious perception resides, unlike the brain region corresponding to the retina or early vision (Block, 2007a, 2011; Coltheart, 1980; Sligte et al., 2008, 2009).

††Research on the relationship between consciousness and attention is a broad topic, see (Block, 2005, 2014a; Graziano, 2022; Koch and Tsuchiya, 2007a,b; Lamme, 2003; Luo and Maunsell, 2019; Maier and Tsuchiya, 2021; Michel et al., 2019)."
always on because whenever he tries to see if it is on, it is always on. Nevertheless, the light actually only comes on when the refrigerator door is open, not vice versa.

According to the fragmentary illusion, subjects only have sporadic or fragmentary consciousness of the (visual) perceptual representation before they receive a tone cue (de Gardelle et al., 2009; Kouider and Dehaene, 2010). Similar to the generic illusion, the tone cue serves to draw the subject’s attention to a particular row, and with the top-down attention it brings, the focused fragmentary consciousness is combined with the representation to form a complete conscious experience. This claim is derived from a variant of the Sperling experiment (de Gardelle et al., 2009), which was designed using slightly different stimuli and reporting procedures. The main difference is that in the additional phase, one letter in the original array is replaced by a rotated or flipped form (hereafter, pseudo-letter) to test the extent to which the subject consciously perceives the representation in the no-cued region. However, the accuracy of their reports in the additional phase was rather low, but interestingly, they would report the pseudo-letter appearing in the uncued region as its original counterpart with a relatively high probability (0.12 and 0.98, respectively), for example, they would claim to see (by select‡‡) an “M”, when in fact it was an “W” that appeared in the uncued region.

Kouider and Dehaene (2010) interpreted the phenomenon that illusory pseudo-letters seen as real letters to mean that there are some common geometric features between pseudo- and real letters that subjects use in their reports to make judgments while excluding other levels of information or, in some cases, by filling in appropriate content (e.g., subjects with stronger top-down expectations about the possible content of a scene may act in this filler role). According to this interpretation, subjects’ pseudo-letter-based geometric information is only fragmented low-level information, and they may only consciously access this sort of fragmentary information of uncued regions and then automatically report the uncued stuff with letters reconstructed by filling, leading to a supposedly rich and complete report, but this supposed perceptual richness is essentially only a post-constructional illusion.

Phenomenal overflow seems to be confronted with a competing interpretation, overflowers usually have two strategies for responding. On the one hand, to show that there is no illusion as claimed

‡‡It is important to note here that, unlike the original Sperling experiment, this modified experiment did not require subjects to take a verbal report, but rather asked them to freely select in a sequence that was presented that contained, but was not limited to, the previously presented content, see (de Gardelle et al., 2009, 570-573).
by the competitors, which is what most anti-overflows have historically claimed; on the other hand, to show that the evidence used by the competitors is false, or at least flawed.

4. Color diversity, generic illusion and overflowed attention

As a representative of the first strategy, Bronfman et al. (2014) used another variant of the Sperling-type paradigm, the color diversity experiment, in which their data can be interpreted by saying that the subject had concrete and complete consciousness of the representations of the uncued regions, rather than merely experiencing a generic illusion. In contrast to the original Sperling’s, the new paradigm also presented subjects with the same array as before, but the difference was that the content of the array was colored and the color diversity changed (high or low) under different conditions, one of the most critical tasks of this experiment was to test whether subjects could report the color diversity of the content of the uncued region even after their attention was engaged by the main task (reporting the letter content in the cued region). The results show that subjects whose attention was maximally engaged by being asked to report the item in a cued row were still able to accurately report the color diversity of the uncued region because the color diversity comes from the conscious perception of specific letter colors in uncued regions.

The results are taken as mild evidence for phenomenal overflow, refuting the illusory non-specific conscious interpretation of the uncued region, but claiming a concrete and non-generic one. Specifically, it can be interpreted that the specific colors of the item(s) in the uncued region were still consciously experienced, albeit perhaps rather briefly, while their attention was occupied by the target task (focusing on the content of the cued region), for some reason, however, this part of the experience ends up in working memory in a generic or gist-like form to be cognitively accessed for reporting. Not coincidentally, Vandenbroucke et al. (2012) also knew from their experimental design that subjects develop consciousness of the representation outside their attentional focus using the Kanizas triangle experiment.

However, a different interpretation follows from Kouider and Dehaene (2010), who argues that subjects could report color diversity not because they experienced a specific item at all, but because

---

As Bronfman et al. (2014, 1401-1402) notes, this is limited to color information, and other types of information (such as complex shapes and motions, etc.) may still require further scientific analysis.
the experience of the uncued region, which is generic or gist-like with the help of attention, over-
flowed, meaning that even though their attention was maximally engaged by the main task, some
overflowed, causing them to form a gist-like experience (color diversity) of the representation in the
uncued region, but not enough to prove that they developed a concrete, non-generic or complete
experience of it.

But is this the case? Not really. This begins with data from classical inattentional blindness ex-
periments (Simons and Chabris, 1999, 1063-1072), in which subjects are told to count the number of
passes a group of players make about a basketball, and after about thirty seconds of starting this task,
a person dressed as a gorilla walks across the scene and the experiment is stopped after a further pe-
riod of time. Miraculously, about 46% (this percentage varied across experimental conditions) of the
subjects who were focused on counting (without anticipation) did not notice the gorilla-suited indi-
vidual, a result that has been taken by someone as evidence that conscious perception is limited. It
is worth noting, however, that even though close to half did not notice, the remaining subjects (54%)
noticed this intruder, and although it is not explicitly accounted for in the article associated with the
experiment whether they saw more detailed information about the intruder (in fact, this question-
ing component did occur in the experimental procedure), it is briefly mentioned in a later, similarly
derived experiment that some subjects did could report some detailed information (exactly about
the color) outside of the main task’s region of attention, and while this was cursorily interpreted as
possibly due to a slightly weaker ability to occupy attention for that main task, they did not clarify
how to define the main task, what it means to be highly or fully occupied and how to measure and
compare it, and until this is clarified, this cursory interpretation pales slightly and leaves room for
debate.

If Kouider et al. want to use overflowed attention as a reason to interpret subjects’ performance
in the color diversity experiment, then they would also have to face the question of why, in both
types of experiments, subjects’ attention was engaged by the main task, why the former overflowed
attention only led to a gist-like (or generic) ability to report on the representational content of the
unattended region, whereas the latter overflowed attention resulted in subjects having a seemingly
stronger ability to report. As a proponent of the role of overflowed attention at work, Kouider could
indeed argue that this depends on differences in the capacity of the overflowed attention, but he would need to present further evidence, in particular, on the basis of the relevant interpretations given by Bronfman et al. Alternatively, if Kouider et al. concede that the subjects’ performance in the inattentential blindness experiment can also be interpreted by overflowed attention, i.e., that 54% of the subjects were able to see gorillas because of what overflowed, then it seems that they need to further interpret why the other remaining subjects’ attention did not overflow, leading them to report seeing gorillas (even if it was only a vague gist-like silhouette of a gorilla or just a generic image different from the players)?

5. Multiple interpretations still have not gone

As an example of the second strategy, in this section, I will show that Kouider’s fragmentary illusionary interpretation of the results of his experiment is unconvincing and that this description, while appealing, is not the only plausible interpretation. As described in Section 2, subjects would have behaved in the experiment by reporting pseudo-letters as original letters in the free-report procedure, Kouider et al. offer a seemingly more plausible interpretation for this performance, namely that the subjects have a fragmentary consciousness of these pseudo-letters (e.g., about their geometric features), but not a specific one. What can be argued here is, is the overflow really unable to interpret this performance of the subjects? This paper suggests that a recent study in neuroscience may help us to give a negative answer, however, its aim is not to increase the original legitimacy of the phenomenal overflow, but rather to increase the openness of Kouider’s experimental data to multiple interpretations.

It is a common phenomenon that most native speakers can read a jumbled text, whether in Chinese or any other language, without difficulty, for example, “Does the human mind read words as a whole?” Imagine that when we read this kind of text, it is difficult for us to read such sentences in the wrong order without expectation, instead, it seems that the brain automatically adds the out-of-sequence order to the normal order presented to the brain, and then allows us to read them in the normal order very smoothly, even without noticing that some of the letters or text content are reversed at all when we quickly scan the text. This phenomenon was raised by Whitney and Grainger.
of Cambridge University and has generated ongoing attention and discussion. A new development in the study of this phenomenon is the localization of a neural mechanism involved in the recognition of letter codes and subsequent lexical decisions (Agrawal et al., 2020). The results of this study show that the LO area involved in vision characterizes each string with a constitutive code, that the string decision must involve some comparison of stored word representations that are located in the so-called visual word form area (VWFA), and that its activation is related to the time of the string decision afterward.

Thanks to the discovery of this neural mechanism, perhaps during the course of Kouider’s experiment, subjects first generated a visual representation of the pseudo-letter, and then the neural code corresponding to this part of the representation was further activated and used to compare it with the representation in the VWFA area, which was reported after the comparison, it is worth noting that in this experiment, this phenomenon persisted even when subjects were fully aware that their perceptions could be misleading, suggesting that there is indeed a mechanism beyond their subjective control (de Gardelle et al., 2009). Furthermore, Kouider et al. experimentally summarized the illusion as resulting from the interaction between perceptual difficulty and strength of priors: “the poorer the evidence, the more the elaboration of the percept will depend on priors (de Gardelle et al., 2009, 576).” The so-called perceptual difficulty, i.e., the visual representation, and the a priori strength, i.e., the extent to which the letter is preserved in the VWFA area. Letter reports benefit from the visual lexical content stored in VWFA, which is high relevance to strong expertise developed through years of reading. It is important to clarify here that a priori strength does not only refer to the high relevance due to strong expertise developed over years of reading, but there may be other factors at play, such as memory, habit, and so on, but at least it is reasonable to understand a priori strength as the content stored in the VWFA area in Kouider’s experiments.

Conceivably, Kouider et al. argue that this mechanism is compatible with their interpretation, and perhaps subjects remain fragmentary consciousness of the representation in the uncued region while the neural code corresponding to that part of the representation undergoes a comparison or matching procedure within the VWFA area, a procedure that leads subjects to eventually report that part of the representation as the corresponding original letter. The point, however in this paper, is
that this mechanism also offers hope for overflows. As we noted above, the reason why Kouider’s interpretation seems more plausible than that of the overflows is that, if the subject already has a complete or non-fragmented consciousness of the representation of the uncued region, according to the overflows, why would he or she report the pseudo-letter as the original letter? This seems counterintuitive. However, overflows can argue that the subject first generated a neural code about the representation in the LO region before reporting it and that this part of the neural code was matched or compared in the VWFA region, which eventually led to the reporting of the original letters. Thus, overflows can argue that even in the face of the subjects’ performance in Kouider’s experiment, their interpretation remains permissible, let alone rejected and that the experimental data themselves remain open to different interpretations.

To reiterate, although Kouider et al. justified the illusory interpretation and criticism of their opponents’ views by the experimental data they designed, the overflows benefit from this neural mechanism by essentially allowing their interpretation to maintain plausibility in the face of the subjects’ performance in Kouider’s experiment. If Kouider wants to argue for his claim of the fragmentary illusion with the data from his experiment, he will have to provide more favorable evidence for why the fragmentary illusion is stronger than the overflow’s interpretation in the face of the same data, simply because the neural mechanism mentioned above also lends plausibility to the phenomenal overflow.

6. Conclusion

In this article, I simulate adherence to the perspective of overflow’s and proceed from there. Firstly, I underscore Kouider’s distinct interpretation of the data from the color diversity experiment. Kouider posits that these data are still compatible with the generic illusion, however, this appears to raise some issues that are worthy of further discussion. Secondly, the validity of the fragmentary illusion is questioned. Recent findings on neural mechanisms suggest that the experimental data Kouider et al. relied on can still be interpreted by overflow and do not serve as credible evidence for their illusory interpretation. Both illusory interpretations are partially flawed or inadequate in refuting the phenomenon overflow.
It is important to note that adopting the overflow position to contest their adversaries does not necessarily imply that overflow has greater plausibility as an interpretation. Rather, it signals that while both sides endeavor to defend their interpretations with their respective empirical data, these data cannot be entirely excluded from the explanatory scope of the opposing perspective. Additionally, both interpretations face internal challenges, as I showed in section 4. Nevertheless, despite the apparent absence of reliable criteria for deducing the best explanation, exploring this debate can still be significant. It could prompt researchers to more precisely define pertinent concepts, prevent them from falling into a stalemate of competing claims, and offer richer evidence for the advancement of consciousness research, such as recent studies on natural kinds of consciousness.

References


