

Discussion paper

The power of our names, faces, and the Self-Reference Effect: is there more than meets the eye?

Clea Desebrock

In mythological tales, our names and facial images are often gifted a quasi-magical power. When psychologists use these self-representations (or even items we simply imagine are ‘me/mine’) as stimuli in experimental tasks, studies have shown that our perception, memory, decision-making, and actions can be enhanced. The phenomenon has been termed the Self-Reference Effect (SRE). Does an underlying ‘self’ mechanism underpin these effects? Or, do the effects arise because the stimuli are simply more rewarding, familiar, or deeply-encoded? Could the empirical treatment of the SRE be echoing a faulty folk-intuition that the self is a unitary entity? This article briefly explores the colourful history of self-representations, the research field of the SRE, and some of its key challenges.

The power of self-representations

“Little does my lady dream Rumpelstiltskin is my name” (Grimm & Grimm, 1812/2014, p. 181)

Self-representations, such as our names and facial images, have long been ascribed a magical, persuasive, even ominous, power in our imaginations. It is a theme that has persisted and endured through folklore, religious texts, and literatures since antiquity. The tale of ‘Rumpelstiltskin’ or ‘The Name of the Supernatural Helper’ (published by the Brothers Grimm in the 19th century; Grimm & Grimm, 2014) has recently been traced as far back as 4000 BC (Graça da Silva & Tehrani, 2016). Today’s fiction writers continue to deploy this

device, as in J. K. Rowling's 'he-who-must-not-be-named' (Rowling, 1997) in the Harry Potter series, for example. Its effectiveness perhaps owes as much to a resonance with our own experience as to mere tradition: there is a *felt sense* of latent potency when somebody utters our name. Indeed, the-power-of-names notion has also been integral to societal living. In the Ancient Near East, to learn of a person's 'true' (as opposed to 'public') name was believed to confer on the discoverer a preternatural ability to control or even destroy the name-owner in real terms (Wilson, 1969).

Our facial images have also been endowed with a certain potency. Various interpretations of the myth of Narcissus warn us to beware of an excessive attendance to (and obsessional love for) the image of one's own face; versions of the myth of Medusa assume that a mirror-image of the face of the person (or creature) can be used to render powerless the owner of the real face; and today, the ubiquitous obsession with the 'selfie' apparently underlines the irresistible pull of the self-face.

Outside of literature and mythology, the power of self-representations is also apparent in everyday settings. Those attempting to persuade, or seeking compliance have long enlisted the technique of using a person's name to alter the mind-set and/or behaviours of that person (such as in cold-calling and sales pitches). Furthermore, the popularity of the 'Selfieccino' (where consumers' own faces are 'printed' onto their coffee), and the highly-successful marketing campaigns of global brands such as Coca Cola, which offered consumers the chance to see their own names on products, arguably demonstrate a harness-able power in self-representations which can be used to modulate consuming behaviours.

Exploring the Self-Reference Effect

When psychologists have explored the influence of these self-representations in experimental tasks by using the participants' own name or face as stimuli, they have found that the

participants' performance on the task is consistently remarkably enhanced (Sui, He, & Humphreys, 2012). Perhaps this is not surprising, however. Names and faces recall physical individuals, and so play a practical role in societal living. From the perspective of the owner, whatever accolade, retribution, constraint, or freedom the representation attracts, directly affects the safety and wellbeing of the actual person. Furthermore, a person has a lifelong relationship with their representations, a deeply encoded elaborate history, conditioned emotional responses, and explicit response biases involved in preserving and managing the integrity of these representations. Yet when researchers have used *non-exclusive* self-representations as stimuli (i.e. those items shared by others, such as personality traits), task performance is likewise enhanced. For example, in a seminal study by Rogers, Kuiper, and Kirker (1977), participants encoded a list of adjectives either self-referentially (by evaluating whether the trait described them), semantically (by comparing the meaning of words), phonemically (by deciding whether words rhymed), or structurally (by evaluating the size of the text), and were then asked to recall as many of the adjectives as they could. Those lists of adjectives that had been encoded using 'self-reference' were far better recalled than in any other condition.

Researchers have also found that the effects of these self-associated entities pervade even the lowest levels of processing. For example, Moray (1959) gave participants a listening task. In one ear, the participants were fed a continuous verbal message, which they repeated aloud, and in the other ear a short list of simple words was presented. The participants' recognition and recall of the items on the short list was then tested. The participants were not able to remember a single item. The content of the unattended message appeared to be completely rejected by the participants' conscious perception. However, if the participant's name was included in this list, this item could exclusively be recalled. It was as though the

self-referential item had been able to ‘break through’ into the participants’ conscious perception (a phenomenon which is held to underlie the ‘Cocktail Party Effect’).

Caveats for research on the Self-Reference Effect

Much research on the SRE has been carried out since these early studies and the effect is robust across diverse types of self-referential stimuli, and levels and stages of our information processing (from attention and perception, to memory, decision-making, and even our actions; Desebrock, Sui, & Spence, 2018). However, the area has not been without its controversy. In particular, it has been argued that in empirical studies of ‘self-referential processing’ it has often erroneously (or, at best, prematurely) been inferred that the cognitive properties of the SRE must be distinct from those of other factors (Gillihan & Farah, 2005). The key question raised is whether the task-performance modulations we observe and attribute to the SRE are actually self-specific, or whether they are simply enhancements in processing that arise from other properties of self-stimuli. For example, we also see enhancements in our responses when task stimuli are highly-familiar, rewarding, carry positive emotional valence, or have more deeply-encoded and semantically-elaborated representations. In most studies, self-reference effects do outdo other kinds of effects, but given our life-long exposure to, and experience of these items, is this not surprising? Self-representative stimuli could exert an influence on our perceptual and cognitive processing simply by virtue of being (e.g.,) *more* rewarding, or *more* familiar. In other words, although the effects of self-representations may be somewhat elevated in our imaginations, is the SRE fundamentally nothing ‘special’? (Gillihan & Farah, 2005).

To meet the criterion for ‘special’ in this sense, self-referential processing should be both functionally and anatomically dissociated from other kinds of processing. That is, the SRE should be reliant on a cognitive structure that exclusively handles self-referential

processing and consists in neural circuitry that is separate from the circuitry supporting e.g. other-person-referential or general-purpose processing. ‘Language’ processing, for example, (which is considered ‘special’) is predominantly supported by perisylvian brain areas, which are not activated during related but non-language-specific processing e.g. sound recognition (Gillihan & Farah, 2005). Without this kind of evidence, it is hard to argue that the SRE is any more than an emergent property of more general-purpose cognitive processing. Furthermore, in our empirical treatment of the SRE, we may be echoing a faulty folk-intuition that the self is a core, fixed entity (Hood, 2012).

We certainly intuit in everyday life that there is a core, fixed, timeless and persisting entity that constitutes our ‘self’ (Hood, 2012). We also sense that this ‘self’ has a location. When researchers have probed adults’ and pre-school children’s intuitions about where their ‘self’ *is*, both groups invariably indicate that it is housed somewhere within the body, most commonly at the front of the head, between the eyes (Starman & Bloom, 2012). Philosophers, however, have argued that there is scant evidence supporting this so-called ‘ego’ or ‘pearl’ view of the self (Hood, 2012). There is more neuroscientific evidence, argues Hood, for Hume’s ‘bundle theory’. For Hume, “[A person] is nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement.” (Hume, 1888/1978, p. 252). In other words, self simply emerges from this dynamic interplay of experiences. It is an imaginary structure superimposed on the ebb and flow of sensations and cognitions that arise from multiple and distributed systems. Such a structure is thought, by many philosophers and psychologists, to be narrative in form, and it is this that generates the illusion of the self as an entity. “Who we are is a story of our self—a constructed narrative that our brain creates” (Hood, 2012, p. xiii).

An empirical answer to a philosophical question?

Could the SRE, then, merely be an emergent property of general-purpose (e.g., reward) cognitive systems? If self is simply a fictitious structure and in functional and neuroanatomical terms there is no core mechanism or central hub, how can this work? Do our brains not need some means to fundamentally distinguish self from other, and organize the processing linked to self- versus other-referential items to keep track of our world?

One empirical step forward would be to extract a pure measure of self-referential processing; that is, removed from confounds such as stimulus familiarity, reward, elaborate historical associations and so on, and see what remains of the SRE. Another would be to identify underlying neural circuitry dedicated to self-referential processing. But how can this be achieved using self-representative items such as one's own face, name, or autobiographical knowledge? (e.g. which are highly familiar and overlearned). The answer is that it cannot.

This problem has plagued historical studies assessing the SRE. That is, until 2012 when a new paradigm was introduced into the literature (Sui et al., 2012). The new procedure was able to remove confounds of familiarity and overlearning, and later was used to pit effects of the SRE directly against factors such as reward and emotional valence. But, how? The key was to use neutral rather than personally-representative stimuli and to instruct participants to simply associate or 'tag' these items with the concept of 'me'. In the paradigm procedure, participants associated geometrical shapes with personal labels (e.g., self—square, friend—circle, stranger—triangle), and then, in the main task, indicated whether the presented shape—label stimuli matched the learnt associations, or if they had been mixed up (e.g., self—triangle, stranger—square). In the 'self' condition, where participants were evaluating whether the 'me' shape and label stimuli matched, it was found that task performance was greatly enhanced.

In follow-up studies, it was also shown that the SRE could be dissociated (at least in part) from confounding factors inherent in the previous research (e.g., emotional valence: by using images of emotional faces instead of personal labels as stimuli). Furthermore, imaging work revealed that the SRE was supported by neural circuitry that was distinct from circuitry supporting other-person-referential processing, which included the medial pre-frontal cortex (mPFC). The authors proposed that when we process self-referential stimuli, a core self-representation is activated in the mPFC which then functionally connects with domain-specific neural circuitry dependent on the task, to form a dedicated self-network. This core self-mechanism ‘binds’ our information processing across levels and stages of processing (see Sui & Gu, 2017), and so this function and the distinct neural signature of the SRE may indeed suggest that there is something special about the cognitive properties of the self.

Self and the Self-Reference Effect

How does this new evidence sit with the argument that our sense of self as an entity is illusory? It is becoming increasingly clear that the SRE cannot simply be attributed to effects inherent in *person-exclusive* self-representations such as people’s own names and faces, or solely to the effects of underlying general-purpose cognitive processing. The evidence suggests that distinct neural circuitry does indeed support self-referential processing. This ‘SRE-specific’ circuitry may alter depending on the task domain, but self-referential processing seems consistently to be underpinned by a common integrating factor: activations of a neural hub located in the mPFC (see Sui & Gu, 2017). So, just as people tend to intuit, perhaps a pivotal core self-representation could plausibly be housed at the front of the head, somewhere between the eyes. The picture emerging is that the SRE may be (at least partly) dissociated from other effects both in terms of output and the neural circuitry it relies on. However, what these characteristics of self-referential processing mean for the status of the

self, and whether ‘self’ can be reduced to a core neural entity are up for debate. (See Letheby & Gerrans [2017] for an interesting discussion, which also integrates evidence from psychedelic experience). Debate in the SRE field continues around empirical questions such as: can self-referential processing be functionally dissociated from other kinds of processing? Does a dedicated core self-mechanism underlie the SRE, which supports the effects of all self-referential stimuli across domains? Does self-referential processing have a neural signature? Discussion of the area here is considerably simplified, and many questions and much conceptual analysis remain for psychologists, neuroscientists and philosophers.

On a final note, when other people activate our self-referential processing (e.g. by using our name), perhaps the felt sense of a real shift in our brain activity and its impact on our behaviours explains the intuition that our self-representations harbour a potency. Perhaps the SRE can illuminate how the power-in-self-representations notion has so colourfully and persistently permeated our myths and storytelling. The SRE may also offer a mechanism for understanding, for example, the influence of self-representations in the context of ‘personalization’ marketing campaigns. From a psychological perspective, we can conclude that there is clearly more to the SRE than meets the eye.

Acknowledgements:

This manuscript is attached to a PsyPAG bursary.

Correspondence:

Clea Desebrock

DPhil candidate

Department of Experimental Psychology

University of Oxford.

References

- Desebrock, C., Sui, J., & Spence, C. (2018). Self-reference in action: arm-movement responses are enhanced in perceptual matching. *Acta Psychologica*, 190, 258–266.
- Gillihan, S.J., & Farah. M.J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin*, 131 (1), 76–97.
- Graça da Silva, S., & Tehrani, J. J. (2016). Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society Open Science*, 3 (1): 150645.
- Grimm, J., & Grimm, W. (2014). *The complete first edition: The original folk and fairy tales of the brothers Grimm* (J. Zipes, Ed. & Trans.). Princeton, NJ: Princeton University Press. (Original work published in 1812).
- Hood, B. (2012) *The self illusion: how the social brain creates identity*. New York, NY: Oxford University Press.
- Hume, D. (1888/1978). *A treatise of human nature*, second edition. (L.A. Selby-Bigge & P.H. Nidditch, eds.). Oxford: Clarendon Press. Retrieved from [http://www.oxfordscholarlyeditions.com].
- Letheby, C., & Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. *Neuroscience of Consciousness*, 2017 (1): nix016.
- Moray, N. (1959) Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11 (1), 56–60.

- Rogers, T.B., Kuiper, N.A., & Kirker, W.S. (1977). Self-Reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 35 (9), 677–688.
- Rowling, J. K. (1997). *Harry Potter and the philosopher's stone*. London: Bloomsbury.
- Starman, C., & Bloom, P. (2012). Windows to the soul: Children and adults see the eyes as the location of the self. *Cognition*, 123, 313–318.
- Sui, J., He, X., & Humphreys, G.W. (2012). Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, 38 (5), 1105–1117.
- Sui, J., & Gu, X. (2017). Self as object: Emerging trends in self research. *Trends in neurosciences*, 40 (11), 643–653.
- Wilson, J. (1969). The god and his unknown name of power. In J.B. Pritchard (ed.), *Ancient near eastern texts relating to the old testament*, third edition (pp. 12–14). Princeton, NJ: Princeton University Press.