## Research Article

# AUTOMATIC STEREOTYPING

Mahzarin R. Banaji[1] and Curtis D. Hardin[2]

[1]*Yale University and* [2]*University of California, Los Angeles*

**Abstract**—*Two experiments tested a form of automatic stereotyping. Subjects saw primes related to gender (e.g.,* mother, father, nurse, doctor*) or neutral with respect to gender (e.g.,* parent, student, person*) followed by target pronouns (stimulus onset asynchrony = 300 ms) that were gender related (e.g.,* she, he*) or neutral (*it, me*) or followed by nonpronouns (*do, all*; Experiment 2 only). In Experiment 1, subjects judged whether each pronoun was male or female. Automatic gender beliefs (stereotypes) were observed in faster responses to pronouns consistent than inconsistent with the gender component of the prime regardless of subjects' awareness of the prime–target relation, and independently of subjects' explicit beliefs about gender stereotypes and language reform. In Experiment 2, automatic stereotyping was obtained even though a gender-irrelevant judgment task (pronoun/not pronoun) was used. Together, these experiments demonstrate that gender information imparted by words can automatically influence judgment, although the strength of such effects may be moderated by judgment task and prime type.*

Based on recent theory and research on the role of unconscious processes in beliefs about social groups (Banaji & Greenwald, 1994; Bargh, 1994; Greenwald & Banaji, 1995), we report two experiments that provide stricter tests than previously conducted of a form of automatic stereotyping. Several recent experiments have demonstrated that stereotyping can occur implicitly, without subjects' conscious awareness of the source or use of stereotypic information in judgment (Banaji & Greenwald, 1995; Banaji, Hardin, & Rothman, 1993; Devine, 1989). In this article, we focus on a particular brand of stereotyping than can occur even when the perceiver retains awareness of the source of influence on judgment, but is unable to readily control the stereotypic response. Stereotyping, like other cognitive processes, consists of both automatic and controlled components, and the particular form of automaticity that is involved (e.g., awareness, intentionality, efficiency, and controllability) has been of recent interest (see Bargh, 1994). In the present experiments, we demonstrate that gender, as lexically coded in English, can operate automatically in judgment, even when the primary (denotative) meaning is not about gender.[1]

The semantic priming procedure is commonly used to examine automatic information processing and, in particular, to reveal the strength of association between two concepts that exists independently of conscious thought. Developed more than 20 years ago, this procedure has led to important discoveries about attention, signal processing, and semantic memory (Meyer & Schevaneveldt, 1971; Neely, 1977; Posner & Snyder, 1975). The first of these tests showed the now well-known effect that response latency to a target word is facilitated to the extent to which a prime word that appears prior to the target word is semantically related to it. In addition, the technique has recently been successfully adapted to demonstrate the operation of automatically activated attitudes or evaluations (Bargh, Chaiken, Govender, & Pratto, 1992; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Perdue & Gurtman, 1990).

Our primary interest lies in beliefs, and for the present research, we adapted the semantic priming procedure to provide a strict test of the extent to which beliefs about gender (i.e., gender stereotypes) operate automatically. Two words were presented in close succession, and the relationship between them was captured by reaction time (RT) to judge the second (target) word. In both experiments, the central empirical question of interest was, what is the influence of the gender code of a prime on speeded judgments of gender-consistent or gender-inconsistent targets? Faster judgments on targets that follow gender-congruent primes than on targets that follow gender-incongruent primes (i.e., gender-based priming) are taken as evidence for the automatic use of gender stereotypes.[2] We also examine related questions, such as (a) the relationship between automatic stereotyping and traditionally used explicit stereotyping measures, (b) the role of awareness of gender as a potential source of influence on performance, (c) the gender relevance of the judgment task, and (d) the gender strength of the primes.

Although variations of the semantic priming procedure have been used in previous research on stereotypes, the experimental procedures of these studies did not adhere to conventional standards for revealing automatic information use.[3] For example, Dovidio, Evans, and Tyler (1986) presented a prime (*black* or *white*) followed by a target (*intelligent* or *lazy*) and asked subjects if the target "could ever be true" of the prime category. They found that white subjects were reliably faster to respond to stereotype-related traits than stereotype-unrelated traits following the primes *white* and *black*. Such experiments

Address correspondence to Mahzarin Banaji, Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520-8205, e-mail: mbanaji@minerva.cis.yale.edu, or Curtis Hardin, Department of Psychology, UCLA, 1282A Franz Hall, Box 951563, Los Angeles, CA 90024-1563, e-mail: hardin@psych.ucla.edu. All materials may be obtained from the authors.

1. These demonstrations, although showing evidence for the automatic use of gender information, should not be taken to imply that seemingly automatic responses can never be controlled. The effects of automatically activated information are controllable under theoretically specified conditions (Bargh, 1994; Blair & Banaji, 1995).

2. The term *stereotype* has been a construct of changing meaning in social psychology, and our use of it is in keeping with recent definitions that reduce it to refer to beliefs about the attributes of social groups (Ashmore & Del Boca, 1981; Greenwald & Banaji, 1995).

3. In experiments in which the currently specified conditions of automaticity were met, the findings addressed the role of automatic evaluation rather than automatic belief (Gaertner & McLaughlin, 1983; Perdue & Gurtman, 1990, Experiment 2).

  

Commenter et discuter **un seul** des deux articles :
- Banaji & Hardin (1996) « Automatic Stereotyping » *Psychological Science, 7*, 136-142.
- Cohen & Dennett (2011) « Consciousness cannot be separated from function », *Trends in Cognitive Sciences, 15,* 358-364.

Vous pouvez, dans votre commentaire vous aider des questions suivantes, mais vous êtes libre de développer tout autre aspect que vous souhaiteriez.

A/ Sur Banaji & Hardin, 1996 :

1- Quelle définition du stéréotype est adoptée ici ?
2- Quels sont les arguments pour dire que le stéréotypage est *automatique* ?
3- Est-ce que « automatique » et « inconscient » sont équivalents ? Peut-on savoir ici si le stéréotypage est inconscient ?
4- Quel est l'intérêt de la seconde expérience ?
5- Est-ce que cette étude prouve que les participants ont automatiquement des comportements sexistes ?
5- Y a-t-il et si oui quelles seraient selon vous les implications en terme d'éthique, d'éducation et de politiques publiques d'une telle étude ?

B/ Sur Cohen & Dennett, 2011 :

1- Quelle sont les définitions d'une « fonction cognitive » et de « l'accès cognitif » adoptées par les auteurs ?
2- Peut-il y avoir une forme de conscience qui ne se manifeste dans aucun comportement observable ?
3- Pouvez-vous imaginer une expérience réelle ou un déficit neuropsychologique réel qui illustrerait l'expérience de pensée de « l'expérience parfaite » imaginée par les auteurs ?
4- Est-ce qu'une science de la conscience est forcément incomplète ?
5- La vision fonctionnaliste de la conscience des auteurs permet-elle d'étudier la conscience animale et la conscience chez l'enfant pré-verbal ?

Comment and discuss **one and only one** of the two following articles:
- Banaji & Hardin (1996) « Automatic Stereotyping » *Psychological Science*, *7*, 136-142.
- Cohen & Dennett (2011) « Consciousness cannot be separated from function », *Trends in Cognitive Sciences*, *15,* 358-364.

In your commentary you can rely on the following questions, but you can freely expound on any other aspect.

A/ Sur Banaji & Hardin, 1996 :

1- How is "stereotype" defined here?
2- What are the evidence in favor of the notion of *automatic* stereotyping?
3- Are "automatic" and "unconscious" equivalent? Can we here know whether stereotyping is unconscious here?
4- What is the import of the second experiment?
5- Does the study prove that participants have automatic sexist behaviors?
5- Are there, and if so what would be the ethical, educational and public policy implications of such a study?

B/ Sur Cohen & Dennett, 2011 :

1- What are the definitions of "cognitive functions" and of "cognitive access" used by the authors?
2- Can there be a form of consciousness that would not translate into any observable behavior?
3- Can you imagine an actual experiment or neuropsychological deficit that would exemplify the thought experiment of the "perfect experiment"?
4- Is a science of consciousness necessarily incomplete?
5- Would the functionalist conception of consciousness here argued for enable the study of consciousness in animals or pre-verbal infants?

# Consciousness cannot be separated from function


Michael A. Cohen[1] and Daniel C. Dennett[2]

[1] Vision Sciences Laboratory, Department of Psychology, William James Hall, Harvard University, Cambridge, MA, 02138, USA
[2] Center for Cognitive Studies, Department of Philosophy, Tufts University, Medford, MA, 02155, USA



**Numerous theories of consciousness hold that there are separate neural correlates of conscious experience and cognitive function, aligning with the assumption that there are 'hard' and 'easy' problems of consciousness. Here, we argue that any neurobiological theory based on an experience/function division cannot be empirically confirmed or falsified and is thus outside the scope of science. A 'perfect experiment' illustrates this point, highlighting the unbreachable boundaries of the scientific study of consciousness. We describe a more nuanced notion of cognitive access that captures personal experience without positing the existence of inaccessible conscious states. Finally, we discuss the criteria necessary for forming and testing a falsifiable theory of consciousness.**


## The hard problem of consciousness is an impossible problem

A goal of neuroscience is to locate the neural correlates of consciousness: the minimal set of neuronal events leading to subjective awareness (see Glossary) [1–3]. Numerous influential theories hold that conscious experience has its own neural underpinnings that can be separated from all cognitive functions (i.e. attention, working memory, language, decision making, motivation etc.). Different theories equate consciousness with different correlates: recurrent activation between cortical areas [4–7], coalitions of 'winning' neurons [8–10], special microactivations distributed throughout the brain [11–13] or activity in the ventral stream [14]. Although the details of these theories vary, they all assert that conscious experience and cognitive functions have distinct neural correlates (Box 1).

This alleged division between experience and function is often mapped onto the distinction between the 'hard' and 'easy' problems of consciousness [3]. Under this view, the hard problem is answering the question of how phenomenal experience arises from physical events in the brain, whereas the easy problems are characterizing the mechanisms supporting cognitive functions. In this article we argue that, from an empirical perspective, the 'hard problem' is actually an impossible problem that inherently isolates consciousness from all current and future avenues of scientific investigation. All theories of consciousness based on the assumption that there are hard and easy problems can never be verified or falsified because it is the products of cognitive functions (i.e. verbal report, button pressing etc.) that allow consciousness to be empirically studied at all. A proper neurobiological theory of consciousness must utilize these functions in order to accurately identify which particular neural activations correlate with conscious awareness.

A motivation behind dissociative theories is the belief that theories associating awareness with access [15–17] cannot explain the richness of phenomenology. In other words, it is claimed that 'phenomenology overflows access' ([7], p. 487): we experience more than can possibly be captured by cognitive mechanisms that are known to have strict limits. Visual attention [18,19], working memory [20,21], dynamic tracking [22,23] and many other such processes have well-established capacity limits. Phenomenology, however, is claimed to have no such limitations. It is thought that when we look out onto the world we do not only see a few attended items; we see the whole world. Thus it is argued that although we are conscious of a variety of inputs we have access to only a small subset of these experiences [4,7,10,13].

Here, we analyze the data used to support the claim that phenomenology overflows access and show how these results can be accounted for under a pure access/functional view of consciousness. We then argue that dissociative theories are inherently unfalsifiable and beyond the scope of science, because inaccessible conscious states are intrinsically off-limits to investigation. With this in mind we end by describing the necessary components of a proper scientific theory of consciousness.

## Evidence supporting the dissociation

What data support the view that consciousness occurs independently of, and can be experimentally dissociated

### Glossary

**Access consciousness**: conscious states that can be reported by virtue of high-level cognitive functions such as memory, attention and decision making.
**Awareness**: the state of perceiving, feeling or experiencing sensations.
**Easy problem of consciousness**: understanding the mechanisms that support relevant functions such as language and attention.
**Hard problem of consciousness**: explaining phenomenal consciousness (e.g. the feeling of 'what it is like' [60]).
**High versus low-level brain regions**: in this context, the distinction between high and low-level brain regions roughly correspond to sensory and non-sensory (functional) regions. More specifically, 'low-level' regions are involved in the processing and discrimination of visual stimuli, whereas 'high-level' regions are involved in attention, language, and decision-making.
**Phenomenal consciousness**: the subjective aspect of experiencing the world (e.g. the experience of seeing the color red).


*Corresponding authors:* Cohen, M.A. (michaelthecohen@gmail.com); Dennett, D.C. (daniel.dennett@tufts.edu).


 1364-6613/$ – see front matter  doi:10.1016/j.tics.2011.06.008 Trends in Cognitive Sciences, August 2011, Vol. 15, No. 8

**Box 1. Example dissociative theories**

The partitioning of conscious experience from cognitive function is common in neurobiological theories of consciousness. Three representative theories are described below.

*Local recurrency.* The best-known theory that embraces the separation between experience and function is the local recurrency theory put forth by Lamme [4,5] and Block [6,7]. According to this theory, visual information is processed in the cortex by an initial feedforward sweep in which representations of motion, color and shape are formed [61,62]. Although representations at this stage can be rather detailed, no conscious experience accompanies this processing. Such experiences only arise as a result of sustained RP between visual areas. However, the experiences that accompany RP are independent of all cognitive functions, especially attention [4]. Indeed, this theory explicitly maintains that local recurrency is the neural correlate of one and only one form of consciousness: phenomenal consciousness [28]. Access consciousness, which comprises functions such as working memory, language production and so on, is achieved when RP extends into the frontal cortex and engages higher-level functions.

*Microconsciousness.* Zeki's theory of microconsciousness states that consciousness is not a unified state but is instead distributed in space and time [11–13]. Rather than emphasizing the flow of information between regions, like the local recurrency theory, this theory focuses on the activation of 'essential nodes' throughout the cortex. Each node represents different bits of information (e.g. color or motion) and the activation of each node generates its own microconsciousness. We have the impression of a unified consciousness because each of these individual representations is bound to others, post-experientially, to form an accessed macroconsciousness [13]. It is at this macrolevel that functions such as language and decision making operate on the distributed experiences and lead to subjective reports. Thus, the micro/macro distinction again dissociates conscious experience and cognitive function [29].

*Coalitions of neurons.* Crick and Koch proposed that consciousness stems from 'winning' coalitions of neurons (sustained activation of a collection of neurons that are dedicated to the processing and representation of a particular stimulus or event) [8]. Under this view, coalitions supporting one representation compete with coalitions supporting other representations [9]. Only after a winning coalition becomes conscious can attention be diverted to it. Oftentimes, only one coalition 'wins' at a time, leading to a relatively tight correlation between consciousness and attention. However, this correlation is not perfect. Koch and colleagues have written extensively about the existence of consciousness without attention [63,64,69,70], recently claiming that consciousness without attention is a form of phenomenal consciousness as described by Block and Lamme [10].

All of these theories have distinctive strengths, and some plausibility, but they also share a fundamental flaw: they posit the existence of conscious states that even the individual him or herself does not realize he or she is having. Highlighting this flaw might provide impetus for revision and improvement of these theories: rejecting the one shared feature of them all and leaving the other features to be sorted out empirically.

from, higher-level functions (i.e. access)? The most frequently cited experiments use Sperling's partial report paradigm [24–27]. After being briefly shown a display of 9–12 letters, participants can only report some of the items through free recall. However, if cued to report a subset of the letters, they can report the entire subset and thus seem to have consciously perceived all of the items. According to dissociative theories, these results demonstrate that although we have access to only a few items we are nonetheless conscious of the identities of them all [4–7].

Although the partial report results are crucial to arguments for dissociating consciousness and function, they can be explained without this separation [27]. Participants can identify cued items because their identities are stored unconsciously until the cue brings them to the focus of attention. Before the cue, participants are conscious only of the few letters they attend to and the impression that there are other items on the display whose identities they do not know (Figure 1). Once the cue is presented, they are able to access an unconscious representation before it decays and successfully recall the letters presented.

Although Sperling's results can be explained by appealing to consciousness without access, this is not the only explanation or the clearest. Indeed a more nuanced notion of access and cognitive function can readily explain both the phenomenology and the results of these experiments.

## A multi-access model

Those who argue for experience without access [4–14, 28,29] emphasize the introspective experience of seeing

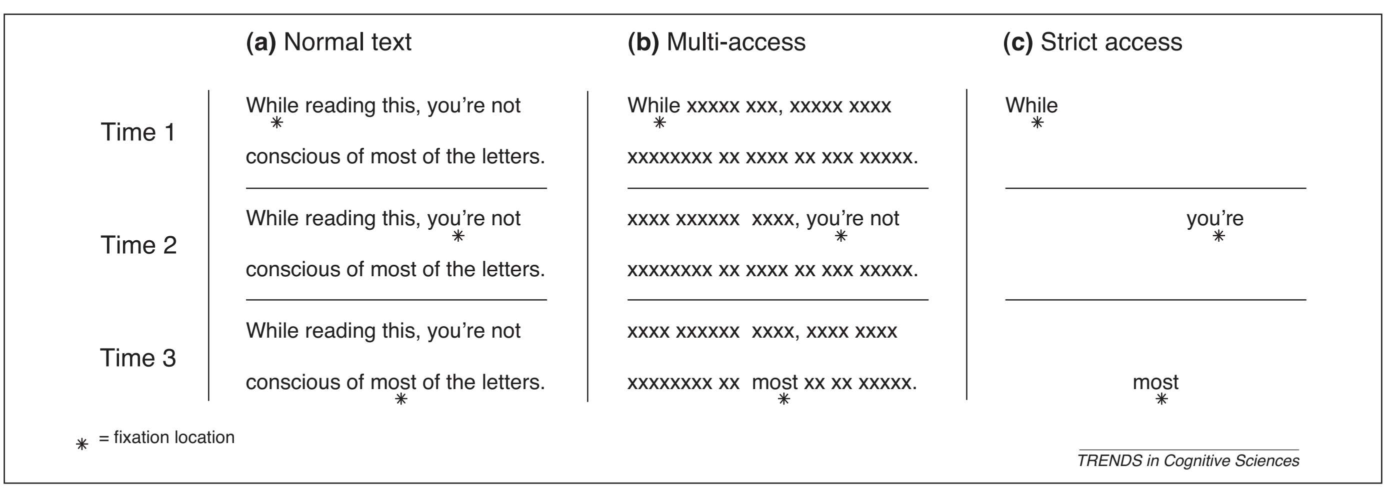


**Figure 1.** How much is consciously perceived at a given psychological moment? Those who argue for dissociating consciousness and function do so because they claim that awareness overflows conscious access. **(a)** In this case, the identities of all the letters on the screen are perceived. **(b)** However, McConkie and Rayner have shown that when uniform Xs replace the nonfixated words of text, participants do not realize this has happened [65,66]. **(c)** If there is no other text on the screen besides the fixated word, then participants will notice this instantly. This elegantly demonstrates that although people are aware of the 'presence' of nonattended items in this case, they are actually not aware of the 'identities' of those items.

a more vivid, detailed world than can be reported. The world beyond focused attention is not in total ‘darkness’: when staring intently at a single item, one is still aware of some aspects of the scene around it [30,31]. Such a claim is obviously true. Dissociative theorists cite this fact as the primary example of phenomenology overflowing access. However, this is not a problem for theories that identify consciousness with function.

The world beyond focal attention is not in darkness because when attention is not entirely engaged by a primary task, and it is unclear if attention can ever be entirely engaged using psychophysical techniques, excess attentional resources are automatically deployed elsewhere [32–35]. Thus, certain items are processed through focal attention, whereas others are processed via distributed attention [36,37]. Focal attention often leads to high resolution percepts whereas the percepts from distributed attention are at a lower resolution but with certain basic elements preserved [36–48] (Figure 2). It is inaccurate to say that information outside the focus of attention receives zero attention. Information not processed by focal attention can nevertheless be the target of other types of attention: distributed, featural, spatial, internal and so on [49].

Is this degraded visual information an example of phenomenology overflowing access? Indeed, the degraded information is consciously perceived. However, the function supporting this perception is simply distributed, rather than focal, attention. In fact, when attention is engaged in a sufficiently difficult task, observers can fail to perceive even coarse and degraded information, such as the gist of a scene, because of inattentional blindness [50]. This information is undoubtedly accessed because observers explicitly report seeing more than what is focally attended (the idea that such information is indeed accessed has been recognized by Block, see [7], p. 487).

Once it is recognized that distributed attention leads to degraded but accessed percepts, the motivation for claiming that this degraded information is an example of inaccessible conscious states disappears. The world beyond focused attention is not in darkness because there are functional resources (in this case, multiple forms of attention) dedicated to processing that information (Figure 1c).

## Is there more to phenomenology?

Dissociative theories claim that there is phenomenology over and above the accessed information previously described. However, various empirical results cast doubt upon this claim. In a modified version of the Sperling paradigm, where letters are sometimes unexpectedly replaced with pseudo-letters, participants still claim to see only letters [51]. Another example of this phenomenon can be seen in Figure 2. When participants are instructed to fixate at the center of a screen, two images can be successively presented in the same location, with a blank image briefly separating the two, and the drastic changes between the images go unnoticed (a phenomenon known as change blindness). If participants are conscious of the identities of all elements in the scene, as has been repeatedly claimed by dissociative theorists, then participants should instantly notice the pseudo-letters or the scrambled image. The fact that they do not suggests that participants are overestimating the contents of their own experience.

Even though people do not notice these changes, the illusion of seeing more still needs to be explained. Why is it that people overestimate the richness of their conscious perceptions [52]? The nature of this illusory experience still needs to be explained and should be the focus of future empirical work. Functionalist accounts can study this by varying the prior expectations and confidence levels of participants in a variety of paradigms [27,51]. Dissociative theories, meanwhile, ‘explain’ this illusion by relying on

TRENDS in Cognitive Sciences

**Figure 2.** Only foveated items are perceived in full color and at high resolution. As stimuli move further into the periphery, they gradually lose their color and fidelity [67–70]. However, although the quality of unattended or unfoveated stimuli is severely degraded, certain basic features and statistics are preserved. In the above example, a natural scene is presented (left) next to an image in which the quality of the image is systematically degraded from the center of the image (the blue fish) towards the periphery. When these two images are presented in rapid succession and with a blank gap in between, observers fixating at the center of the image are unable to detect the differences between the images and claim that they are identical. Observers do not notice that a single isolated percept is so degraded because they are able to move their eyes throughout a scene with so little effort that this behavior is often overlooked [31].

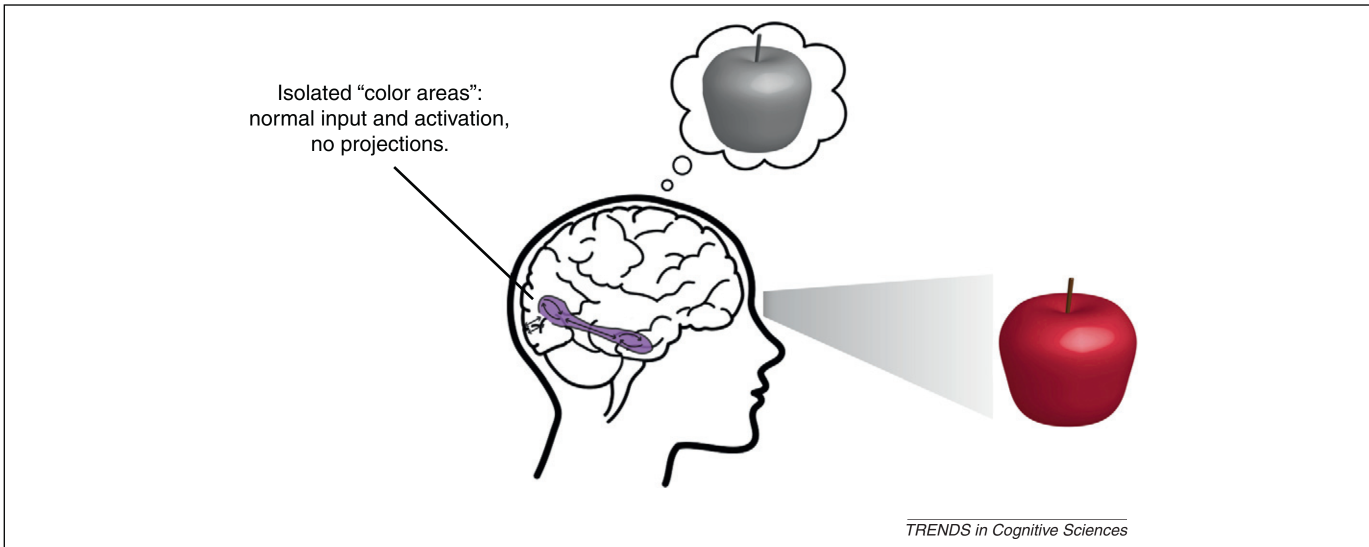


**Figure 3**. A graphic depiction of the perfect experiment. When presented with a red apple there will be normal activation of the color areas of the brain but without projections to higher–level areas. Other areas of the brain (e.g. object representation and identification, language production etc.) will function normally, so the patient will be able to report that he or she sees an apple but an apple that has no colors.

inaccessible conscious states that, as the next section describes, inherently prevent the possibility of confirmation or falsification.

## The perfect experiment

Currently, no experimental results uniquely support the existence of consciousness independent of function and access. Could future experiments accomplish this? We argue that all theories of consciousness that are not based on functions and access [4–14] are not scientific theories. Consider perhaps the most drastic experiment possible, the 'perfect' experiment: imagine that, in the future, surgeons are able to isolate the parts of the visual cortex that represent color while wholly preserving their activation patterns. After this surgery, the areas involved in color perception (visual area V4, inferotemporal cortex etc.) behave normally but are simply unable to project to higher brain areas [53–57]: perfect isolation. Although the color areas are isolated, all other visual areas (e.g. motion, luminance, object recognition etc.) are untouched and project to higher-level regions in a normal manner (Figure 3). Such a clean separation of one aspect, color, of visual perception is profoundly unrealistic but this idealization provides a simplification that is revealing of the key flaw in theories that dissociate function from consciousness.

According to all the theories discussed above, or possible theories based on the experience/function divide, whatever is necessary for color consciousness will be preserved in these color areas. If these theories are mutually exclusive, then we can imagine a different participant for each particular theory. All that matters is that we do not allow these isolated areas of a supposed type of phenomenal consciousness to interact with other cognitive functions.

When shown a colored apple what will our hypothetical participants say? They will surely not say that they see any colors because the areas responsible for processing color have been isolated from higher-level areas, including language production. They will be able to identify the object as an apple because visual areas responsible for all other aspects of visual cognition are intact and connected to these higher-level regions. Thus, they are simply colorblind. We can imagine them saying, 'I know you say my color areas are activated in a unique way, and I know you believe this means I am consciously experiencing color but I'm looking at the apple, I'm focused on it, and I'm just not having any experience of color whatsoever' (Box 2).

Moreover, imagine that, before the surgery, that particular shade of red would reliably agitate or excite the patient. Would the patient have such feelings now and say something like, 'I don't see red but I notice that I've gotten a little tense'? As described here, the patient would not because such affective, emotional or 'limbic' reactions are themselves the types of functions that we are isolating from the color area. To be excited or calmed or distracted by a perceptual state of red discrimination is already to have functional access to that state, however coarse-grained or

**Box 2. What if we gave the isolated color area the ability to communicate?**

Is it possible that even though the subject is not conscious of red, the isolated color area itself is experiencing color (similar to the way the right hemisphere of split-brain patients is often described) [11]? What would happen if we supplied a reporting mechanism for the isolated color area?

Imagine the reporting mechanism is nothing more than the simple hardware needed to actually transmit a message (e.g. a speaker). If this device were connected to the color area, then it seems clear that there would be no reports of color consciousness. The cognitive functions needed to select a particular thought, decide how best to describe it, and to execute that action are still absent, preventing any type of response from being formed or conveyed.

Whereas if the color area were connected to a more sophisticated reporting mechanism that was endowed with these functions there would probably be reports of color consciousness. However, this is not because the color area is experiencing its own isolated consciousness; rather, it is because the color area is now connected to the functions that are crucial for consciousness. By connecting the color area to a mechanism endowed with the relevant functions, the previously unconscious color information can now be accessed by a broader cognitive system.

**Box 3. Access when there is no behavioral output**

How does the relationship between access, function and consciousness apply if a person cannot move or give any type of behavioral response? Consider patients with locked-in syndrome. Patients with this condition are conscious but cannot move due to paralysis of all muscles except (usually) the eyes and eyelids. This small volitional movement is the only means by which they can communicate. Imagine, however, that even this behavior is disabled so the patient is still fully conscious but completely paralyzed: perfectly locked-in.

This is an important case for understanding the functional view: behavioral outcomes are not its defining component. People can still consciously experience the world without there ever being any behavioral result that follows from those experiences. What is important is that there are enough high-level functions engaged with that information such that the patient could volitionally act upon those experiences if he or she so desired and was not paralyzed. In this case, even though the patient cannot move, the patient can do things such as attend to what he or she is hearing or store selected bits of information in working memory. This patient being conscious is perfectly consistent with the functional account of consciousness because those functions are fully preserved.

incomplete, because such a reaction can obviously affect decision making or motivation (Box 3).

In spite of this frank denial by subjects, theories that posit dissociation between consciousness and function would necessarily assume that participants of the 'perfect experiment' are conscious of the apple's color but simply cannot access that experience. After all, the conditions these theories stipulate for phenomenal consciousness of color are all met, so this experiment does not disprove the existence of isolated consciousness; it merely provides another particularly crisp example of consciousness without access.

However, there is a crucial problem with this logic. If this 'perfect experiment' could not definitively disprove dissociative theories, then what could? The subject manifests all the functional criteria for not being conscious of color so what would ground the claim that the subject nevertheless enjoys a special kind of consciousness: phenomenal consciousness without access consciousness (Box 2)?

## The domain of a science of consciousness

What the perfect experiment demonstrates is that science necessarily relies on cognitive functions in order to investigate consciousness. Without input from subjects, input that is the product of such functions, theorists are left to define consciousness based on certain types of activation that are independent of a subject's own experience. It has been claimed that separating consciousness from other cognitive functions is required because it 'is a prerequisite for using the term [consciousness] at all' ([5], p. 500).

What does it mean to study consciousness without function? Inevitably, theories motivated by this view will define consciousness in their own way (local recurrency, microconsciousness, coalitions of neurons, etc.) and say that whenever that criterion is met, consciousness must occur. But how do we set this criterion?

For example, what reason is there to think that local recurrency is conscious experience? Could local recurrency simply be a form of unconscious processing? It cannot be based on subjective reports because these reports are the direct result of cognitive functions. When an observer says, 'But in the Sperling display I don't just see a few letters on the screen, I see all the letters,' there is no reason to believe that such an experience occurs independent of function.

The fact that the observer is reporting on this visual experience proves that the experience has been accessed by the broader cognitive system as a whole. Lamme writes, 'You cannot know whether you have a conscious experience without resorting to cognitive functions such as attention, memory or inner speech' ([5], p. 499). If this is true, then what reason is there to think this particular type of activation should be classified as correlating with conscious experience? What does it mean to have a conscious experience that you yourself do not realize you are having? In the face of such clear grounds for doubting such a conscious experience, dissociative theories need to provide a reason for claiming that these isolated types of activation involve any kind of consciousness.

## The future of scientific theories of consciousness

It is clear, then, that proper scientific theories of consciousness are those that specify which functions are necessary for consciousness to arise. A true scientific theory will say how functions such as attention, working memory and decision making interact and come together to form a conscious experience. Any such theory will need to have clear and testable predictions that can in principle be verified or falsified. Most importantly, such theories will not claim that consciousness is a unique brain state that occurs independently of function; instead, the focus will be placed on the functions themselves and how they interact and come together to form consciousness.

There are several theorists who have already realized the need for functions in developing theories of consciousness. Dehaene and colleagues [16] have put forth a global neuronal workspace model that claims consciousness is defined by the orientation of top-down attention, long-distance feedback loops that extend into parietofrontal networks, and conscious reportability. Similarly, Kouider and colleagues [27] have discussed at great length how information that is in consciousness relies on a hierarchy of representational levels. Under this view, each level corresponds to different cognitive mechanisms responsible for different units of representation.

It is important to stress that both of these theories are merely the beginning, rather than the end, of the study of consciousness. There is still much work to be done in regards to how these functions and mechanisms interact. In Dehaene *et al*.'s theory, for example, a more thorough and specific understanding of the type of parietofrontal activation [16] and how it relates to the formation of memories and decisions is still necessary. The upshot of function-based theories is that they make claims about consciousness that can be tested and examined scientifically.

Although there are certainly those who disagree with the specifics of the theories put forth by Dehaene *et al*. and Kouider *et al*. [4–14], these are disagreements that can eventually be settled through more rigorous examination and testing. The same cannot be said of theories that maintaining that consciousness occurs independent of function. As the perfect experiment illustrates, such theories inherently prevent any future avenue for scientific research.

## Concluding remarks

Understanding the necessary relation between function and experience reveals that the so-called hard problem of consciousness should be reclassified. Far from being a formidable obstacle to science, it achieves its apparent hardness by being systematically outside of science, not only today's science but any science of the future that insists on dissociating consciousness from the set of phenomena that alone could shed light on it. This is not to suggest that consciousness is a mystery that the human mind cannot comprehend [58]. It is simply that whatever mysteries and puzzles might continue to baffle us, we should not cripple our attempts at understanding by adopting a concept of consciousness that systematically blocks all avenues of further research.

The issues raised here generalize beyond the specific theories discussed [4–14]. Any theory wherein the neural correlates of conscious experience are separate from the neural correlates of cognitive function is ultimately doomed. No matter the specifics of the theory – C-fibers firing, grandmother cells, winning coalitions, microconsciousness, recurrent processing (RP) and so on – it is always possible in principle to isolate this activation. Such imagined isolation, however, actually removes the experience in question from further testing, scrutiny and verification. Although these theories might provide considerable insight into the formation of internal representations of the sensory and perceptual world, that is not enough to explain one's personal awareness. A proper theory of consciousness cannot exclusively focus on how the brain forms and maintains representations. Such a theory must also explain in functional terms how those representations are experienced and accessed by the multiple functions constituting an observer [59]. Theories that do not acknowledge this are fundamentally incapable of explaining the full scope of consciousness.

### Acknowledgments

This work was supported by a National Science Foundation Graduate Research Fellowship (M.A.C.). Special thanks to Justin Jungé for extensive discussion and comments on the project. Thanks to Arash Afraz and Maryam Vaziri Pashkam for helpful discussions, and to Ray Jackendoff, Sid Kouider, Ken Nakayama, Jordan Suchow, and two anonymous reviewers for comments on the manuscript. Thanks to Jeremy Freeman for providing the images for Figure 2.

### References

1 Crick, F. and Koch, C. (1995) Are we aware of neural activity in primary visual cortex. *Nature* 375, 121–123
2 Crick, F. and Koch, C. (1990) Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275
3 Chalmers, D.J. (2000) What is a neural correlate of consciousness? In *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (Metzinger, T., ed.), pp. 17–39, MIT Press
4 Lamme, V.A. (2003) Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18
5 Lamme, V.A. (2006) Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501
6 Block, N. (2005) Two neural correlates of consciousness. *Trends Cogn. Sci.* 9, 46–52
7 Block, N. (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–499
8 Crick, F. and Koch, C. (2003) A framework for consciousness. *Nat. Neurosci.* 6, 119–126
9 Koch, C. (2004) *The Quest for Consciousness: A Neurobiological Approach*, Roberts & Company
10 Koch, C. and Tsuchiya, N. (2007) Phenomenology without conscious access is a form of consciousness without top-down attention. *Behav. Brain Sci.* 30, 509–510
11 Zeki, S. (2003) The disunity of consciousness. *Trends Cogn. Sci.* 7, 214–218
12 Zeki, S. and Bartels, A. (1999) Toward a theory of visual consciousness. *Conscious. Cogn.* 8, 225–259
13 Zeki, S. (2001) Localization and globalization in conscious vision. *Annu. Rev. Neurosci.* 24, 57–86
14 Goodale, M. (2007) Duplex vision: separate cortical pathways for conscious perception and the control of action. In *The Blackwell Companion to Consciousness* (Velmans, M. and Schneider, S., eds), pp. 580–588, Blackwell
15 Baars, B.J. (1989) *A Cognitive Theory of Consciousness*, Cambridge Univ. Press
16 Dehaene, S. *et al.* (2006) Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* 10, 204–211
17 Dennett, D.C. (1991) *Consciousness Explained*, Little Brown
18 Awh, E. and Pashler, H. (2000) Evidence for split attentional foci. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 834–846
19 McMains, S.A. and Somers, D.C. (2004) Multiple spotlights of attentional selection in human visual cortex. *Neuron* 42, 677–686
20 Luck, S.J. and Vogel, E.K. (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281
21 Alvarez, G.A. and Cavanagh, P. (2004) The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol. Sci.* 15, 106–111
22 Pylyshyn, Z.W. and Storm, R.W. (1988) Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 179–197
23 Cavanagh, P. and Alvarez, G.A. (2005) Tracking multiple targets with multifocal attention. *Trends Cogn. Sci.* 9, 349–354
24 Sperling, G. (1960) The information available in brief visual presentation. *Psychol. Monogr.* 74, 1–29
25 Landman, R. *et al.* (2003) Large capacity storage of integrated objects before change blindness. *Vision Res.* 43, 149–164
26 Sligte, I. *et al.* (2008) Are there multiple visual short-term memory stores? *PLoS ONE* 3, 1–9
27 Kouider, S. *et al.* (2010) How rich is consciousness? The partial awareness hypothesis. *Trends Cogn. Sci.* 14, 301–307
28 Block, N. (1995) On a confusion about the function of consciousness. *Behav. Brain Sci.* 18, 227–287
29 Zeki, S. (2007) A theory of microconsciousness. In *The Blackwell Companion to Consciousness* (Velmans, M. and Schneider, S., eds), pp. 580–588, Blackwell
30 Wolfe, J.M. (1999) Inattentional amnesia. In *Fleeting Memories: Cognition of Brief Visual Stimuli* (Coltheart, M., ed.), pp. 71–94, MIT Press
31 O'Regan, J.K. and Noë, A. (2001) A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–1031
32 Yi, D.J. *et al.* (2004) Neural fate of ignored stimuli: dissociable effects of perceptual and working memory load. *Nat. Neurosci.* 7, 992–996
33 Lavie, N. (1995) Perceptual load as a necessary condition for selective attention. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 451–468
34 Cartwright-Finch, U. and Lavie, N. (2006) The role of perceptual load in inattentional blindness. *Cognition* 102, 321–340
35 Schwartz, S. *et al.* (2005) Attentional load and sensory competition in human vision: modulation of fMRI responses by load at fixation during task-irrelevant stimulation in the peripheral visual field. *Cereb. Cortex* 15, 770–786
36 Chong, S.C. and Treisman, A. (2003) Representation of statistical properties. *Vision Res.* 43, 393–404
37 Alvarez, G.A. (2011) Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* 15, 122–131
38 Oliva, A. and Schyns, P.S. (2000) Diagnostic colors mediate scene recognition. *Cogn. Psychol.* 41, 176–210
39 Alvarez, G.A. and Oliva, A. (2008) The representation of ensemble visual features outside the focus of attention. *Psychol. Sci.* 19, 678–685

40 Alvarez, G. and Oliva, A. (2009) Spatial ensemble statistics: efficient codes that can be represented with reduced attention. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7345–7350
41 Ariely, D. (2001) Seeing sets: representation by statistical properties. *Psychol. Sci.* 12, 157–162
42 Chong, S.C. and Treisman, A. (2005) Statistical processing: computing the average size in perceptual groups. *Vision Res.* 45, 891–900
43 Halberda, J. *et al.* (2006) Multiple spatially overlapping sets can be enumerated in parallel. *Psychol. Sci.* 17, 572–576
44 Dakin, S. and Watt, R. (1997) The computation of orientation statistics from visual texture. *Vision Res.* 37, 3181–3192
45 Parkes, L. *et al.* (2001) Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* 4, 739–744
46 Atchley, P. and Andersen, G. (1995) Discrimination of speed distributions: sensitivity to statistical properties. *Vision Res.* 35, 3131–3144
47 Watamaniuk, S. and Duchon, A. (1992) The human visual system averages speed information. *Vision Res.* 32, 931–941
48 Haberman, J. and Whitney, D. (2007) Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* 17, R751–R753
49 Chun, M.M. *et al.* (2011) A taxonomy of external and internal attention. *Annu. Rev. Psychol.* 62, 73–101
50 Cohen, M.A. *et al*. Natural scene perception requires attention. *Psychol. Sci.* (in press)
51 De Gardelle, V. *et al.* (2009) Perceptual illusions in brief visual presentations. *Conscious. Cogn.* 18, 569–577
52 Levin, D.T. *et al.* (2000) Change blindness: the metacognitive error of overestimating change-detection ability. *Vis. Cogn.* 7, 397–412
53 Pascual-Leone, A. *et al.* (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–512
54 Haynes, J. *et al.* (2005) Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron* 46, 811–821
55 Boehler, C.N. *et al.* (2008) Rapid recurrent processing gates awareness in primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8742–8747
56 Super, H. *et al.* (2001) Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nat. Neurosci.* 4, 304–310
57 Bullier, J. (2001) Feedback connections and conscious vision. *Trends Cogn. Sci.* 5, 369–370
58 McGinn, C. (1999) *The Mysterious Flame: Conscious Minds in a Material World*, Basic Books
59 Jackendoff, R. (1987) *Consciousness and the Computational Mind*, MIT Press
60 Nagel, T. (1974) What is it like to be a bat? *Philos. Rev.* 4, 435–450
61 Lamme, V.A. and Roelfsema, P.R. (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579
62 Lamme (2000) Neural mechanisms of visual awareness: a linking proposition. *Brain Mind* 1, 385–406
63 Koch, C. and Tsuchiya, N. (2007) Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci.* 11, 16–22
64 Tononi, G. and Koch, C. (2008) The neural correlates of consciousness – an update. *Ann. N. Y. Acad. Sci.* 1124, 239–261
65 McConkie, G.W. and Rayner, K. (1975) The span of effective stimulus during a fixation in reading. *Percept. Psychophys.* 17, 578–586
66 Rayner, K. (1975) The perceptual span and peripheral cues in reading. *Cogn. Psychol.* 7, 65–81
67 Newton, J.R. and Eskew, R.T., Jr (2003) Chromatic detection and discrimination in the periphery: a postreceptoral loss of color sensitivity. *Vis. Neurosci.* 20, 511–521
68 Hecht, E. (1987) *Optics*, Addison Wesley
69 Wandell, B.A. (1995) *Foundations of Vision*, Sinauer
70 Azzopardi, P. and Cowey, A. (1993) Preferential representation of the fovea in the primary visual cortex. *Nature* 361, 719–721

were critical in setting the stage for the present study, but did not strictly test automatic stereotyping. Most notably, the time between the onset of the prime and target (stimulus onset asynchrony, or SOA) in the previous studies was long enough to allow strategic processing, casting doubt about the automaticity of the process being measured. In the present experiments, we used a 300-ms SOA, a condition known to capture relatively automatic processes (Neely, 1977, 1991).

Further, Dovidio et al. (1986) required subjects to deliberately link the prime and target by asking whether the target "could ever be true" of the prime category, and Neely (1977) demonstrated that such explicit expectations do affect RT under long SOAs such as those used by Dovidio et al. (1986). Although these judgment tasks demonstrate important differences in judgment latencies for stereotyped traits following the prime *black* or *white*, these tasks do not index automatic processes that may occur outside conscious deliberation of prime–target relationships. In the present experiments, we used judgment tasks that required no attention to the relationship between prime and target. Indeed, subjects were instructed to ignore the prime word and classify the target word as a male or female pronoun (Experiment 1) or a pronoun or not a pronoun (Experiment 2).

In addition, the experiments we report differ from previous research in the number and type of stimuli that were used. Instead of the repeated presentation of two-category labels as primes (*black* or *white*, *young* or *old*), we used 150 primes signifying gender in a variety of ways, including those associated to gender by definition (e.g., *mother, father, waiter, waitress*) or by normative base rates (e.g., *doctor, nurse, mechanic, secretary*), or neutral with respect to gender (e.g., *humanity, citizen, people, cousin*). The larger set of primes more fully represents the social category, allows use of primes other than category labels alone, and permits a comparison of the strength of primes that denote gender and of primes that connote gender. Primes also included so-called generic masculine terms (e.g., *mankind*) to allow a test of whether such words automatically connote maleness or perform the more inclusive function that critics of nonsexist language assert is the case.

In the choice of target words, we departed from the almost exclusive reliance of past research on trait adjectives. In both experiments, we used pronouns because they inescapably mark gender (e.g., *she, he*). However, the judgment task itself differed across the two experiments in whether the decision focused on gender (male or female; Experiment 1) or grammatical form (pronoun or not pronoun; Experiment 2). To date, no studies of stereotyping have used a task that does not focus attention on the category of interest (e.g., gender, race). A finding that automatic stereotyping occurs even when a gender-irrelevant task is used would attest to the potency of automatic gender stereotypes.

## EXPERIMENT 1

Experiment 1 tested whether gender information in words is automatically used in judgment as assessed by faster response times when the genders of the prime and target words match (e.g., *doctor-he, nurse-she*) than mismatch (e.g., *doctor-she, nurse-he*). In addition, measures of beliefs about explicit gender stereotypes, language reform, and the influence of gender in everyday life were included to test the relationship between automatic stereotyping and more traditional explicit measures of gender stereotyping.

### Method

#### *Subjects*

Sixty-eight subjects (32 female, 36 male) from the introductory psychology pool at Yale University participated in partial fulfillment of a course requirement.

#### *Materials and apparatus*

The experimental task was administered on IBM-PS2 microcomputers running Micro-Experimental Laboratory software (Schneider, 1990). Subjects entered judgments on protruding keys, marked "M" and "F," affixed to the *f* and *j* keys. Key position was reversed for half the subjects.

Two hundred primes were divided evenly among four categories: male related, female related, neutral with respect to gender, and nonword letter string (*ZZZZZ*). Within each of the first three prime categories, words were chosen to appear virtually equally from two subcategories. The first subcategory contained words associated to gender by normative base rates. These words were chosen on the basis of 1990 census data indicating occupations that were heavily skewed (over 90%) toward the participation of either females (e.g., *nurse, secretary*) or males (e.g., *doctor, mechanic*) or that had equal participation (e.g., *reporter, postal clerk*). In addition, several other words having strong stereotypical associations to one gender or the other were included (e.g., *feminist, god*). The second subcategory contained words associated to gender by definition, that is, words that expressly refer to gender (e.g., *woman, man*), kinship terms (e.g., *mother, father*), and titles (e.g., *mr, mrs, king, queen*). Within this subcategory, words containing male morphemes (e.g., *salesman*), female morphemes (e.g., *salesgirl*), or neutral morphemes (e.g., *chairperson*) were also included. Targets were the six most common pronouns in English, half male (*he, him, his*) and half female (*she, her, hers*).

Three measures were designed to assess explicit beliefs regarding gender stereotypes, language reform, and the influence of gender in peoples' lives.[4]

#### *Design and procedure*

For each trial, events occurred in the following order: First, an orientation symbol (+) appeared for 500 ms. Then the prime word appeared for 200 ms, followed by a blank screen for 100 ms. Finally, the target pronoun appeared and remained on the screen until a response was entered. Subjects made 432 judgments (not including practice and buffer trials) divided equally among the eight prime-target categories (prime: male, female,

4. For a description of these measures, see Hardin and Banaji (in press).

neutral, nonword; target: male pronoun, female pronoun) within each of three blocks of trials that were counterbalanced across subjects. Prime and target stimuli were paired randomly for each subject.

The design was a 4 (prime gender: female, male, neutral, nonword) × 2 (target gender: female, male) × 2 (subject gender: female, male) mixed factorial, with subject gender the between-subjects factor. Subjects judged each pronoun as either male or female. They were instructed to ignore the primes and judge the targets as quickly and accurately as possible. Subjects then completed the three explicit measures of gender beliefs. Finally, they were probed for their awareness of the hypotheses and debriefed.

### Results and Discussion

Reported results are based on correct judgments, excluding responses that were extreme outliers. Consistent with other studies employing this procedure, the error rate was low (1,117 of 29,502 judgments, or 3.8%). RTs greater than 3 *SD* above the mean (>1,300 ms) were identified as outliers and excluded (208 trials, or 0.7%). In sum, 95.6% (28,193 judgments) were retained in the reported analyses. The pattern of results is unchanged when these data are included. To achieve a better approximation to the normal distribution, analyses were performed on a log transformation of the raw RT latencies. Thirty-seven of 68 subjects were aware of the gender relationship between primes and targets. However, consistent with the assumption that this procedure reflects relatively automatic processing, the pattern of results was identical for both aware and unaware subjects, all $F$s $< 1$.

As shown in Figure 1, the predicted gender priming effect was obtained, indicating that judgment was faster when target gender matched than mismatched prime gender. The omnibus Prime Gender (female, male, neutral, nonword) × Target Gender (female, male) × Subject Gender (female, male) three-way analysis of variance yielded the predicted Prime Gender × Target Gender interaction, $F(3, 198) = 72.25$, $p < .0001$. The specific Prime Gender × Target Gender interaction (excluding the

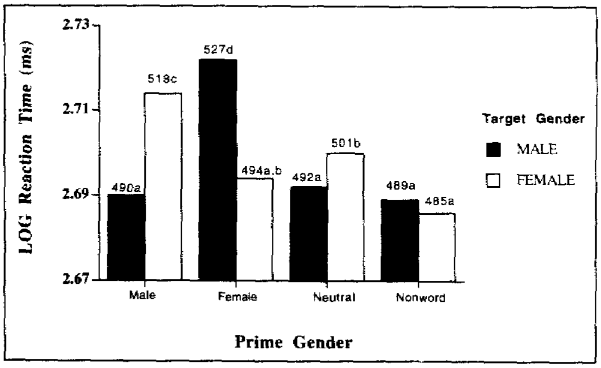


Fig. 1. Mean reaction time to judge words as male or female as a function of prime gender and target gender (Experiment 1, $n = 68$). Bars with shared subscripts are not significantly different from each other ($p > .05$).

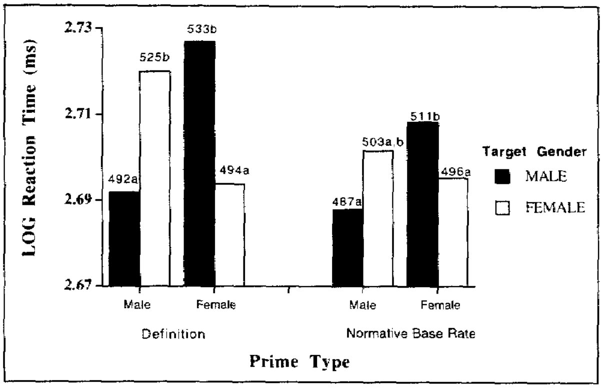


Fig. 2. Mean reaction time to judge words as male or female as a function of prime type, prime gender, and target gender (Experiment 1, $n = 68$). Bars with shared subscripts are not significantly different from each other ($p > .05$).

neutral conditions) was also reliable, $F(1, 66) = 117.56$, $p < .0001$. Subjects were faster to judge male pronouns after male than female primes, $t(67) = 11.59$, $p = .0001$, but faster to judge female pronouns after female than male primes, $t(67) = 6.90$, $p = .0001$. In addition, subjects were faster to respond to targets preceded by male ($M = 2.702$) than female primes ($M = 2.708$), $F(1, 66) = 7.52$, $p < .01$. No other reliable main effects or interactions were obtained as a function of either subject gender or target gender ($F$s $< 1$).

The automatic gender priming effect was obtained for primes related to gender both by definition (e.g., *mother, father, man, woman*), $F(1, 67) = 103.97$, $p < .0001$, and by normative base rate, $F(1, 67) = 18.61$, $p < .0001$. However, as shown in Figure 2, the gender priming effect was significantly larger for primes related to gender by definition, as revealed by the three-way Prime Type (definition, normative base rate) × Prime Gender (female, male) × Target Gender (female, male) interaction, $F(1, 67) = 13.67$, $p < .0005$.

Generic masculine terms contributed to the automatic gender priming effect. After primes containing the morpheme *man* (e.g., *fireman, mankind, human, man*), judgments were faster for male pronouns ($M = 2.687$) than female pronouns ($M = 2.709$), $t(67) = 4.06$, $p < .0001$. The relationship also held under the most conservative analysis, in which terms that are sometimes used to refer only to men (e.g., *man, fireman*) were excluded. Primes considered to be generic masculine terms in virtually all contexts (e.g., *mankind, layman*) produced faster judgments for male pronouns ($M = 2.689$) than female pronouns ($M = 2.712$), $t(67) = 2.18$, $p < .05$.

Finally, we examined terms that differed in no way except for the gender of their suffix (e.g., *chairman, chairwoman, chairperson*). As expected, the gender of the suffix did influence response latencies as indicated by a Prime Gender (male, female, neutral) × Target Gender (male, female) interaction, $F(2, 112) = 11.59$, $p < .0001$. Judgments were faster for male pronouns after words with male ($M = 2.693$) than female ($M = 2.722$) suffixes, $t(67) = 3.29$, $p < .01$, whereas judgments were marginally faster for female pronouns after primes with female ($M = 2.694$) than male ($M = 2.716$) suffixes, $t(67) = 1.81$, $p =$

.07. In addition, judgments were faster when primes with female suffixes were followed by female pronouns ($M$ = 2.694) than male pronouns ($M$ = 2.722), $t(67) = 2.82$, $p < .01$. Interestingly, neutral *–person* suffixes after the identical words did not produce equivalent responses to female and male pronouns. Instead, after these primes, subjects were still faster to judge male targets ($M$ = 2.704) than female targets ($M$ = 2.722), $t(67) = 2.64$, $p < .05$.[5]

Relations between explicit beliefs and automatic gender stereotyping were examined by computing a correlation between each of the three explicit belief measures and a gender priming score, which was calculated by subtracting log RT for gender-congruent trials from log RT for gender-incongruent trials. None of the three correlations of explicit measures with the priming score was significant (language reform: $r[67] = -.003$, $p = .978$; role of gender in everyday life: $r[68] = -.050$, $p = .686$; explicit gender stereotypes; $r[66] = .037$, $p = .767$). This result is consistent with other research demonstrating a lack of correspondence between explicit and implicit measures of stereotyping (Banaji & Greenwald, 1995).

In sum, Experiment 1 provided evidence for automatic gender stereotyping using a broad range of primes and using time and task parameters that reflect automatic information use. The effect occurred regardless of subjects' awareness of the prime–target relation, and independently of explicit beliefs about gender stereotypes. The effect was also obtained for both primes related to gender by definition and primes related to gender by normative base rate, although not surprisingly the effect was larger for primes related to gender by definition.

## EXPERIMENT 2

Participants in Experiment 1 judged whether each target was male or female, thereby focusing attention on the gender of the target. This form of the judgment task is quite conventional. For example, when theoretical interest has focused on the semantic link between prime and target, the commonly used judgment task is a lexical decision (word/nonword; Neely, 1991). Likewise, when the interest is in the evaluative component of the prime and target, the task is typically a good/bad judgment (Bargh et al., 1992; Fazio et al., 1986; Greenwald, Klinger, & Liu, 1989; Perdue & Gurtman, 1990). However, stronger evidence for automaticity may be obtained if the effect is observed when the judgment task is unrelated to the dimension of the prime–target relationship. For example, Bargh, Chaiken, Raymond, and Hymes (in press) showed that the automatic evaluative effect is obtained even when the judgment involves mere pronunciation, a task unrelated to evaluation. Hence, in Experiment 2, the judgment task was a pronoun/not pronoun decision, unrelated to gender.

### Method

#### *Subjects*

Sixty subjects (29 female, 31 male) from Yale University participated in exchange for $5 or in partial fulfillment of a course requirement.

#### *Materials, design, and procedure*

For this experiment, 120 of the primes used in Experiment 1, representing male (40 primes), female (40 primes), and neutral (40 primes) categories, were selected. Of the four target pronouns used, *she* and *he* allowed the comparisons of primary interest. The pronoun *it* was included because it is the most frequently occurring gender-neutral pronoun, and *me* was included for exploratory purposes to examine a possible relationship between prime gender and subject gender (cf. Markus, 1977). The four nonpronouns (*is, do, as, all*) were chosen to match the critical targets in length, number of syllables, and frequency (Kuçera & Francis, 1967).

In all, each subject made 720 experimental judgments divided into five blocks of trials, counterbalanced across subjects. For 480 of these judgments, the correct response to the question "Is this a pronoun?" was "yes," and for 240, the correct answer was "no." Each prime was paired with (a) both critical "yes"-response targets (i.e., *she, he*), (b) both noncritical "yes"-response targets (i.e., *it, me*), and (c) two of the four "no"-response targets (i.e., *do, all, is, as*). For each subject within each block, prime and target items were randomly associated. After completing the priming task, subjects were probed for awareness regarding the hypotheses and debriefed.

### Results and Discussion

As before, results are based on a log transformation of the raw RT latencies for correct judgments, excluding outliers (RT > 1,300 ms or > 3 $SD$ above the mean; 1.4% of the total). Also as in Experiment 1, the error rate was low (370 of 28,800 "yes" judgments, or 1.3%; 928 of 43,200 total judgments, or 2.1%); 97.7% (28,134) of the "yes" judgments were retained in the reported analyses. Seven of the 60 subjects revealed some knowledge of a possible gender relation between the prime and target words. Again, however, the pattern of results was identical for both aware and unaware subjects, but no statistical significance tests were conducted because of the small number of aware subjects.

As Figure 3 shows, the predicted gender priming effect was obtained, indicating that judgment was faster when target gender matched than mismatched prime gender. The omnibus 3 (prime, gender: male, female, neutral) × 4 (target gender: *she, he, it, me*) × 2 (subject gender: male, female) analysis of variance yielded the predicted Prime Gender × Target Gender interaction, $F(6, 336) = 3.66$, $p < .01$. In addition, a Subject Gender × Target Gender interaction indicated that subjects were faster to respond to targets that matched rather than mismatched their own gender, $F(3, 168) = 3.58$, $p < .02$.[6] No difference in responding to the male and female targets was observed for dubiously neutral primes such as *layman* and *man-*

5. However, we found (Hardin & Banaji, in press) no bias favoring males in a similar experiment using first names as targets.

6. A similar finding was reported by Zarate and Smith (1990). In addition, a main effect of prime gender indicated that subjects' responses were fastest following male primes and slowest following female primes, $F(2, 112) = 3.89$, $p < .03$. A main effect of target gender indicated that subjects were slower to respond to the target *it* than to *she, he,* and *me*, $F(3, 168) = 154.33$, $p < .0001$.

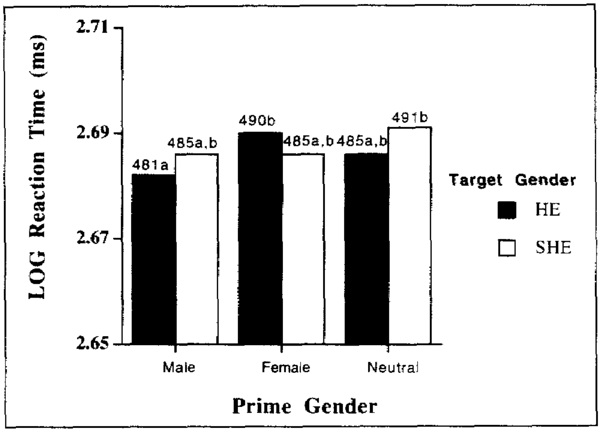


**Fig. 3.** Mean reaction time to judge words as pronouns or not pronouns as a function of prime gender and target gender (Experiment 2, $n = 58$). Bars with shared subscripts are not significantly different from each other ($p > .05$).

*kind*. There was also no main effect of subject's sex, $F(1, 56) = 1.49$.

The more specific interaction of prime gender by target gender (excluding neutral primes) was also significant, indicating that subjects were faster to judge targets in gender-congruent prime-target pairs than in gender-incongruent pairs, $F(1, 56) = 4.63$, $p < .04$. Again, the Subject Gender × Target Gender interaction was reliable, indicating that subjects were faster to respond to the target pronouns that were consistent than inconsistent with their own social category, $F(1, 56) = 17.95$, $p < .0001$.

However, these 2 two-way interactions were qualified by a three-way Subject Gender × Prime Gender × Target Gender interaction, $F(1, 56) = 4.15$, $p < .05$. For purposes of clarity, we describe results separately for primes related to gender by definition and primes related to gender by normative base rates. Analyses of primes related to gender by definition (e.g., *mother, father, waitress, waiter*) yielded the gender priming effect unmoderated by subject gender (Fig. 4, left panel). RT was smaller when prime gender and target gender were congruent than incongruent, as indicated by a reliable two-way interaction, $F(1, 56) = 8.70$, $p < .005$. Subjects were faster to identify *he* when primes were male than female, $t(59) = 2.44$, $p < .02$, but faster to identify *she* than *he* when the primes were female, $t(59) = 2.53$, $p < .02$. In addition, subjects were faster to identify targets that matched their own gender, as indicated by the reliable interaction between subject gender and target gender, $F(1, 56) = 7.43$, $p < .01$.

Analyses of primes related to gender by normative base rates (e.g., *secretary, mechanic, doctor, nurse*) suggest limitations to the generality of automatic gender priming under conditions in which the task does not require subjects to focus on the dimension of gender (see Fig. 4, right panel). Although reliable effects were obtained for the Subject Gender × Target Gender interaction, $F(1, 56) = 14.02$, $p < .0001$, and there was a main effect of prime gender, $F(1, 56) = 6.49$, $p = .01$, both were qualified by a marginal three-way Subject Gender × Prime Gender × dTarget Gender interaction, $F(1, 56) = 3.49$, $p < .07$. Male subjects were faster to identify *he* than *she* regardless of prime gender, $F(1, 29) = 7.44$, $p = .01$, and faster to identify targets after male than female primes, $F(1, 29) = 9.21$, $p < .01$. In contrast, female subjects were faster to identify *she* than *he* after female primes, $t(28) = 2.81$, $p < .01$, and faster to identify *he* after male than female primes, $t(28) = 2.18$, $p < .05$. Females were also faster, in general, to identify *she* than *he*, $F(1, 27) = 6.70$, $p < .05$.

## GENERAL DISCUSSION

These two experiments provide the first strict tests of a form of automatic stereotyping. Using a large number and wide range of stimuli, we demonstrated that judgments of targets that follow gender-congruent primes are made faster than judgments of targets that follow gender-incongruent primes. This effect was obtained despite subjects' deliberate attempt to ignore the prime, regardless of whether subjects were aware or unaware of the gender relation of prime-target pairings, independently of subjects' explicit beliefs about gender, regardless of whether the judgment was gender relevant or irrelevant, and on both words that are gender related by definition and words that are gender related by normative base rates.

The results, however, also show two moderators of the gender priming effect. First, the effect was stronger when the judgment was gender relevant (e.g., male or female pronoun?) than gender irrelevant (e.g., pronoun or not pronoun?). Further research will investigate whether this difference also obtains on other forms of gender-irrelevant tasks, such as pronunciation. Second, the gender priming effect was stronger for primes related to gender by definition (e.g., *mother, father*) than by normative base rate (e.g., *doctor, nurse*). In Experiment 1, for example, the effect size for definition primes was large (Cohen's $d = .78$), whereas for normative-base-rate primes, the effect size was moderate ($d = .47$). This difference reflects the differential strength of the two types of primes in evoking gender. Words that are exclusively reserved to denote gender will produce stronger priming than words that connote gender (for a replication with names as targets, see Hardin & Banaji, in press).

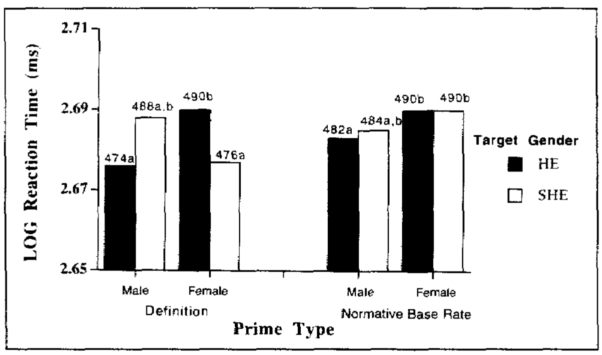


**Fig. 4.** Mean reaction time to judge words as pronouns or not pronouns as a function of prime type, prime gender, and target gender (Experiment 2, $n = 58$). Bars with shared subscripts are not significantly different from each other ($p > .05$).

Sapir (1963) commented that one of the important functions of language is to repeatedly declare to society the psychological status of its members. These experiments show the automatic effects of such repeated linguistic declarations, in particular, those that convey the social psychological positions that are occupied through gender. A noteworthy aspect of the gender priming effect observed in Experiment 1 is that the effect can obtain not only when the primes denote gender (*man*, *woman*), but also when they more tacitly connote gender (*mechanic, nurse*). That gender-signifying information permeates thought sufficiently to influence judgment points to the fundamental nature of gender as a category in verbally communicated thought. This article is not the place to catalogue the various ways in which gender is coded in most languages, but we note that English stands out as one language that has received a quite extensive analysis of what might be called the "genitalia of language" (Baron, 1986): gender-signifying words, gender-specific pronouns, and the covert presence of gender in grammatical structure. We expect that such automatic gender priming effects are best observed in languages that provide extensive and deep coding of gender in grammar and semantics (Hardin & Banaji, 1993). Further evidence for the generality of automatic gender-stereotyping effects might be obtained by demonstrating such effects independently of language (i.e., through the use of nonverbal, pictoral stimuli that denote and connote gender). Such effects would be especially important in revealing the degree to which the present effect is a function of gendered language per se or gender stereotypes more generally.

Although research on beliefs and attitudes has usually depended on direct, verbal measures of stereotypes (see Greenwald & Banaji, 1995), response latencies may provide a more indirect measure of stereotype strength. A case for RT as a measure of attitude or evaluation has already been effectively made (see Bargh et al., 1992; Fazio et al., 1986; Perdue & Gurtman, 1990), and other investigators have used RT as an indicator of stereotypes (Dovidio et al., 1986). However, these experiments, in conjunction with others (Blair & Banaji, 1995; Hardin & Banaji, in press), demonstrate the operation of beliefs under conditions that meet currently accepted standards for measuring automatic processes. Such measures are likely to increasingly complement the more traditional measures of evaluation and belief, especially as their validity and feasibility are further established.

**Acknowledgments**—This research was supported in part by Grants DBC 9120987 and SBR 9422291 from the National Science Foundation. We are grateful to Lisa Driscoll and John Beauvais for assistance with data collection; Kimberly Hinds and Irene Blair for assistance with programming; and R. Bhaskar, Irene Blair, Richard Hackman, and Eliot Smith for comments on a previous draft.

## REFERENCES

Ashmore, D.D., & Del Boca, F.K. (1981). Conceptual approaches to stereotypes and stereotyping. In D.L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–36). Hillsdale, NJ: Erlbaum.

Banaji, M.R., & Greenwald, A.G. (1994). Implicit stereotypes and prejudice. In M.P. Zanna & J.M. Olson (Eds.), *The psychology of prejudice: The Ontario Symposium* (Vol. 7, pp. 55–76). Hillsdale, NJ: Erlbaum.

Banaji, M.R., & Greenwald, A.G. (1995). Implicit stereotyping in false fame judgments. *Journal of Personality and Social Psychology, 68*, 181–198.

Banaji, M.R., Hardin, C., & Rothman, A.J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology, 65*, 272–281.

Bargh, J.A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R.J. Wyer & T. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J.A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic activation effect. *Journal of Personality and Social Psychology, 62*, 893–912.

Bargh, J.A., Chaiken, S., Raymond, P., & Hymes, C. (in press). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology*.

Baron, D. (1986). *Grammar and gender*. New Haven, CT: Yale University Press.

Blair, I.V., & Banaji, M.R. (1995). *Automatic and controlled processes in gender stereotyping*. Unpublished manuscript, Yale University, New Haven, CT.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Dovidio, J.F., Evans, N., & Tyler, R.B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22–37.

Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., & Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.

Gaertner, S.L., & McLaughlin, J.P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46*, 23–30.

Greenwald, A.G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4–27.

Greenwald, A.G., Klinger, M.R., & Liu, T.J. (1989). Unconscious processing of dichoptically masked words. *Memory and Cognition, 17*, 35–47.

Hardin, C., & Banaji, M.R. (1993). The influence of language on thought. *Social Cognition, 11*, 277–308.

Hardin, C., & Banaji, M.R. (in press). Gender in language and thought. *Social Cognition*.

Kuçera, N., & Francis, W.N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35*, 63–78.

Meyer, D., & Schevaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.

Neely, J.H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Perdue, C., & Gurtman, M. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology, 26*, 199–216.

Posner, M.I., & Snyder, C.R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), *Information processing in cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Sapir, E. (1963). *Selected writings of Edward Sapir in language, culture, and personality* (D. Mandelbaum, Ed.). Berkeley: University of California Press.

Schneider, W. (1990). *MEL users guide: Computer techniques for real time psychological experimentation*. Pittsburgh: Psychology Software Tools.

Zarate, M.A., & Smith, E.R. (1990). Person categorization and stereotyping. *Social Cognition, 8*, 161–185.

(Received 12/8/94; Accepted 3/26/95)