Acta Psychologica 203 (2020) 103009

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Involuntary attentional shifts as a function of set and processing fluency

Katelyn Gardner[a], Erica B. Walker[a], Yanming Li[a], Adam Gazzaley[b,c], Ezequiel Morsella[a,b,*]

[a] Department of Psychology, San Francisco State University, United States of America
[b] Department of Neurology, University of California, San Francisco, United States of America
[c] Departments of Psychiatry and Physiology, University of California, San Francisco, United States of America

ARTICLE INFO

Keywords:
Reflexive imagery task
Involuntary attention
Selective attention
Unconscious processing
Mental imagery

ABSTRACT

In laboratory tasks, involuntary cognitions of various kinds (e.g., mental imagery) have been elicited by external stimuli. These effects reveal, among other things, the capacities of involuntary processes. In most cases, these cognitions do not require, for their generation, executive functions such as a shift in selective attention. In Experiment 1, subjects were presented with a clock of 12 words in the stead of numbers and were instructed to focus on the center of the screen and to not count the number of letters of a word at a certain location. Involuntary counting of the critical word occurred on 39% of the trials. This effect requires an involuntary shift of attention. Experiment 2, involving Chinese ideographs, concerned the effect of stimulus fidelity and processing fluency. Native English speakers and a separate group of subjects who could read Chinese ideographs were presented with an array similar to that of Experiment 1 and instructed to not read any of the words. Some words were easy to read (e.g., regular Chinese words and English words), and some words were more difficult to read (e.g., Chinese "loan" words and English pseudowords). For the subjects who could read Chinese ideographs, more involuntary reading occurred for regular ideographs than for loan words. For the Native English speakers, comparable effects were found with the English stimuli. Together, these studies reveal that attentional phenomena of this kind can be influenced involuntarily and systematically through external control.

Most of the contents that compose the *conscious field*[1] arise effortlessly, passively, and involuntarily (Morsella, Godwin, Jantz, Krieger, & Gazzaley, 2016a). For example, after an unexpected nap, the eyes open and one immediately experiences percepts and urges—the sight of a white coffee mug, the smell of coffee, and the urge to change one's posture. To the observer, these *conscious contents*[1] simply "just happen" (Morsella et al., 2016a). Most percepts enter consciousness in this way (Allen, Krisst, Montemayor, & Morsella, 2016; Firestone & Scholl, 2016). Research in perception reveals that entry into consciousness of this nature ("involuntary entry," for short) is influenced by many variables, including the salience, novelty, motion, or incentive/emotional quality of the stimulus (Gazzaley & D'Esposito, 2007; Goodhew, 2017).[2] It is important to note that urges, too, can enter consciousness in this way (Loewenstein, 1996; Morsella, et al., 2009; Morsella, Gray, Krieger, & Bargh, 2009). Investigations on action control have illuminated that involuntary entry of urges can arise from bodily needs (Loewenstein, 1996) and from the activation of conflicting action plans (Desender, van Opstal, & van den Bussche, 2014; Lewin, 1935; Morsella, Gray, et al., 2009; Morsella, Wilson, et al., 2009; Questienne, Atas, Burle, & Gevers, 2018).[3]

Other forms of involuntary entry stem from a combination of sets[4] and external stimuli. For example, as noted by Ach (1905/1951), if one has activated the set to add before hearing "two and two," then one will experience the conscious content "four." If, however, the set was not to add but to subtract, then one would experience "zero" instead of "four." This form of involuntary entry has been referred to as *set-based entry*

* Corresponding author at: Department of Psychology, San Francisco State University, 1600 Holloway Avenue, EP 301, San Francisco, CA 94132-4168, United States of America.
E-mail address: morsella@sfsu.edu (E. Morsella).

[1] Each thing one is conscious of is referred to as a "conscious content" (e.g., a yellow afterimage or nausea). The *conscious field* is composed of all the conscious contents activated at one time.

[2] The mechanisms underlying the involuntary entry of contents into consciousness seem to vary across modalities. For example, though a "pop-out" effect (Treisman & Gelade, 1980) may influence entry in vision, it is less likely to do so in olfaction.

[3] Experimental manipulations have revealed that metacognitions (e.g., action-related urges) can enter consciousness systematically and insuppressibly as a function of set and the presentation of external stimuli (Garcia, Bhangal, Velasquez, Geisler, & Morsella, 2016).

[4] Sets, such as mindsets or task sets, are dispositions to behave or think in certain ways.

https://doi.org/10.1016/j.actpsy.2020.103009
Received 20 April 2019; Received in revised form 6 December 2019; Accepted 8 January 2020
0001-6918/

(Bhangal, Merrick, Cho, & Morsella, 2018), which has been contrasted with the involuntary entry mentioned above concerning percepts (e.g., a bright stimulus) and visceral urges (e.g., thirst).

1. The reflexive imagery task

The Reflexive Imagery Task (RIT; Allen, Wilkins, Gazzaley, & Morsella, 2013; see Review in Bhangal, Cho, Geisler, & Morsella, 2016) was developed to investigate the nature of involuntary entry from a combination of external stimuli and set. The task stems from the instructions of the classic flanker task (Eriksen & Eriksen, 1974), in which subjects must respond to a target stimulus and ignore distractors, but are nonetheless influenced by the distractors in several ways. Specifically, the RIT stems from "subjective" variants of the Eriksen flanker task (e.g., Morsella, Gray, et al., 2009, Morsella, Wilson, et al., 2009; see discussion in Desender et al., 2014, and in Questienne et al., 2018) in which distractors activate involuntary urges and other conscious contents.[5] Other aspects of the task are based on theoretical developments (Morsella et al., 2016a; see Discussion) and on the experimental work by Ach (1905/1951), Stroop (1935), Uznadze (1966), Wegner (1989), and Gollwitzer (1999).

In the task, subjects are instructed to not perform a mental operation (e.g., to count) on to-be-presented stimuli. For example, before being presented with three triangles, subjects might be instructed to not count the number of objects presented on the screen (Bhangal et al., 2018), or, before being presented with the line drawing of a cat, subjects might be instructed to not think of the name of the to-be-presented visual object (Allen et al., 2013). On most trials, despite the intentions of the subject, the undesired mental operations still arise, yielding "three" for the triangles and "cat" (i.e., /k/, /œ/, and /t/) for the stimulus CAT. Part of the effect stems from sets being activated somehow by the negative instructions. The paradigm uses negative instructions only in order to diminish artifacts stemming from demand characteristics and strategic processing on the part of the subject. However, without such negative instructions, RIT effects still arise at comparable rates.[6] The set-based entry in the RIT effect is believed to be involuntary and to reflect what usually occurs in everyday life, when entry into consciousness "just happens."

To illustrate the most basic version of the RIT effect, we will present momentarily to you, the reader, a stimulus object enclosed within parentheses. Your task is to *not* subvocalize (i.e., 'say in one's head') the name of the object. Here is the stimulus (▲). When presented with these instructions (which induce a certain set) and then presented with this stimulus, most people cannot suppress the conscious experience of the phonological form of the word "triangle."

RIT effects of a more complex nature have been obtained by Merrick, Farnia, Jantz, Gazzaley, and Morsella (2015) and by Cho, Zarolia, Gazzaley, and Morsella (2016). In Merrick et al. (2015), subjects were presented with line drawings of objects and instructed to (a) not think of the name of the object, and (b) not count the number of letters composing the object name. RIT effects arose for both mental operations on a significant proportion of the trials (~30%). In Cho et al. (2016), subjects first learned to transform words according to a rule resembling that of the childhood game of Pig Latin. After training, subjects were presented with words and instructed to not transform the words according to the newly-learned rule. Involuntary transformations arose on a substantive proportion of trials (~40%). It is worth noting that this involuntary effect requires, not only memory retrieval, but also symbol manipulation, a process associated with frontal cortex (B. L. Miller & Cummings, 2007).

2. The validity of subjects' self-reports

The self-reports stemming from an RIT can be inaccurate as a result of many factors, including demand characteristics and inaccurate memories of ephemeral conscious contents (Block, 2007). An example regarding the former would be subjects basing their self-reports on knowledge of how one should conduct oneself in an experiment (see discussion in Morsella, Wilson, et al., 2009). However, there is some behavioral evidence that subjects' self-reports are accurate.

First, in Cushing, Gazzaley, and Morsella (2017), subjects reported the occurrence of the basic RIT effect but, in addition, they had to press a button if the involuntary subvocalization they experienced rhymed with a word held in mind. Performance (> 80% mean accuracy across trials) provided evidence that subjects in an RIT experiment do experience involuntary subvocalizations, for detecting a rhyme requires retrieval of the phonological form of a word. Second, in Bhangal et al. (2018), an RIT that was motivated by the research by Ach (1905), subjects were presented with a visual array of objects and instructed to not count the number of objects presented on the screen. Subjects indicated if they involuntarily counted the number of objects and reported the sum. When the number of objects was small (2–5 objects), the counting was very accurate (~90% mean accuracy). This degree of accuracy suggests that the counting did in fact occur as reported by the subject. Third, in RITs in which subjects are instructed to not think of the name of the to-be-presented object, involuntary subvocalizations are influenced by the word frequency of the name of the object: High-frequency words are more likely to yield an RIT effect than low-frequency words, and the latency of the effect is shorter for the former (Bhangal, Merrick, & Morsella, 2015). Such a frequency effect would be unlikely to stem from strategic processing or demand characteristics, for it would require for subjects to know how a variable such as word frequency should influence the nature of responses. Regarding the possibility of the RIT effect arising from strategic processing, we should add that, on many trials, the effect arises too quickly to be caused by such processing (Allen et al., 2013; Cho, Godwin, Geisler, & Morsella, 2014). In such trials, it is unlikely that the involuntary effect arises from subjects having long mentations such as, "I should not think of the name of the object, which is *X*."[7] In addition, some neuroimaging data (which did not stem from RITs) corroborate that subjects do not confabulate about their reported mental events (Mason et al., 2007; McVay & Kane, 2010; Mitchell et al., 2007; Pasley et al., 2012; Wyland, Kelley, Macrae, Gordon, & Heatherton, 2003).

[5] The flanker task precedes research on *ironic processing* (see Footnote 8), which is associated with failures of self-regulation (e.g., in dieting; Wegner, 1989). (The ironic effect was noted long ago by Dostoevsky, 1863.) It should be mentioned that the RIT, unlike research on ironic processing, was designed to investigate, not failures in self-regulation, but the nature of involuntary entry from sets and external stimuli. In short, research on the RIT and on ironic processing stem from different theoretical backgrounds and are concerned with different phenomena and with answering different questions.

[6] It is important to note that RIT effects have arisen in RITs that lack any kind of negative instruction to not perform some kind of mental operation (e.g., see the Baseline Condition in Allen et al., 2013). To take one example, in Allen et al. (2016), subjects were instructed to hold in mind, for as long as possible, one way of perceiving an ambiguous object (e.g., Necker cube). Although there were no negative instructions, involuntary "perceptual reversals," which involve involuntary entry into consciousness, occurred on around 80% of the trials.

[7] Additional behavioral data that corroborate subjects' reports about the RIT effect are the following. The RIT effect still arises when there is cognitive load, a condition in which it is difficult for subjects to implement any form of strategic processing (Cho et al., 2014). Last, RIT effects are systematically more likely for some sensory modalities than for others. For example, RIT effects are more likely for visual imagery and verbal imagery than for olfactory imagery or gustatory imagery (Dou, Li, Geisler, & Morsella, 2018). Such a systematic pattern of results is unlikely to arise from strategic processing or demand characteristics.

3. The involuntary nature of the RIT effect

The notion that the RIT effect is involuntary stems not only from subjects reporting their inability to thwart the effect (despite using a plethora of strategies; Cho et al., 2014) and from behavioral data (revealing, for example, that the effect could not stem from strategic processing), but also from theoretical explanations of the effect. For example, according to Wegner (1994), ironic effects[8] such as the RIT effect arise from the operations of a "monitoring" process that brings into consciousness representations that conflict with intended goals. In the account by Wegner (1994), this monitoring process is usually unconscious, autonomous, and requires little mental effort. In other, perhaps more parsimonious accounts of the RIT effect (Ach, 1905/1951; Bhangal et al., 2016), the effect is the consequence of sets being activated incidentally by the instructions. From this standpoint, merely hearing the word "add" in the instruction "Do not add the following numbers" increases the activation level of the set to add, which thereby yields "four" in response to the stimuli "2 and 2." Having activated the set to subtract in this way would have yielded "zero" in response to the very same stimuli. This account is consistent with the tenets of *parallel distributed processing* (Rumelhart, McClelland, and the PDP Research Group, 1986). Of most importance, in every theoretical account of the RIT effect, including those involving cross-modal imagery (see discussion in Dou et al., 2018), the nature of the effect is involuntary.

4. Boundary conditions of the RIT effect and their theoretical importance

It is worth noting that the RIT, with its focus on involuntary processing, provides an additional method with which to contrast the capacities of conscious and unconscious processes. It is a method that tests some of the limits of involuntary processes without relying on subliminal stimuli, which can be problematic: These imperceptible stimuli are not only unconscious, but they are also of very weak strength, unlike the supraliminal stimuli that unconscious mechanisms often process (Bargh & Morsella, 2008). (The notion that "unconscious = subliminal" led to an underestimation of the sophistication of the unconscious processes which operate during everyday circumstances; see discussion in Bargh & Morsella, 2008.) It could be said that the RIT involves the Helmholtzian-Freudian unconscious, which operates over supraliminal stimuli (as in the case of Helmholtz's, 1856/1925 *unconscious inference*; see related account in Nisbett & Wilson, 1977). With this in mind, it is important to consider that, in a pilot study (n = 8, trials = 8; see Acknowledgment), no RIT effects were observed when the stimuli (orthographs) were rendered subliminal through visual masking.

Identifying the boundary conditions of the RIT effect could illuminate the kinds of processes that might not be able to unfold unconsciously. In turn, this could help isolate the kinds of processes that require volitional, conscious mediation. As mentioned above, RITs have yielded null effects when the stimuli are subliminal. In addition, RITs have yielded null findings when the effect involves emotional processing[9] or basic processes associated with autonomic function (e.g., an RIT task in which the instruction is to not dilate one's pupils; Bhangal et al., 2016). The null effects associated with autonomic function (e.g., the pupillary reflex) have led to the hypothesis that the RIT effect is associated with a subset of the activities of the corticospinal tract (Morsella et al., 2016a), a hypothesis that is consistent with the more general view that conscious processing is in the service of the somatic nervous system (Morsella, 2005; Morsella et al., 2016a).

Of import, the RIT effect is not likely to arise for overt action. In the theorizing that led to the development of the RIT (e.g., Bargh & Morsella, 2008; Morsella, 2005; Morsella et al., 2016a; Morsella, Godwin, Jantz, Krieger, & Gazzaley, 2016b), there is the distinction between the suppressibility of overt behavior and of the generation of conscious contents: Although one could easily suppress the expression of an action plan, one cannot so easily suppress the consciously experienced inclinations (e.g., action-related urges) associated with that action plan. For instance, when fasting, one can suppress the act of reaching for food more easily than suppress the desire to eat food. As Bargh and Morsella (2008) note, inclinations can often be behaviorally suppressible but not mentally suppressible, a difference that is proposed to be adaptive in ontogeny (see discussion in Morsella et al., 2016a, 2016b). Hence, the suppression involved in the RIT is different in nature from the suppression of overt action. Accordingly, in the initial RIT study (Allen et al., 2013), subjects were instructed to (a) not think of the name of the visual stimulus, and (b) not utter the name of the visual stimulus. The RIT effect involving subvocalization occurred often across the trials (~85%), but there was never a trial in which the subject involuntarily uttered the name of the object. According to Allen et al. (2013), this is consistent with the view that the suppressibility of behavior is different from that of the generation of conscious contents.

One concern about the RIT effect in Allen et al. (2013), which involved the involuntary subvocalization of object names, is that the effect is not noteworthy because stimulus-elicited memory retrieval is often automatic (Schacter & Tulving, 1994). This led to the hypothesis that, outside the domains of autonomic function and emotional processing, RIT effects should not arise for mental phenomena requiring symbol manipulation. However, such a hypothesis cannot account for the RIT effects found in Cho et al. (2016), in which the involuntary verbal imagery required symbol manipulation, which is more than just memory retrieval. However, one could argue that even the effect in Cho et al. (2016), too, is not noteworthy, because it is well known that syntactic operations are unconscious even though they involve symbol manipulation.

5. Extension of the RIT effect to the realm of selective attention

Perhaps the true boundary conditions of RIT effects (outside the domains of emotion and autonomic function) lie in mental operations requiring executive processes such as set-based shifts in selective attention. "Selective attention," which has been contrasted with phenomena such as the visual grasp reflex (Sumner & Husain, 2008), has been defined as "the skill through which a person focuses on one input or one task while ignoring other stimuli that are also on the scene" (Reisberg, 2015, p. 611). No RIT to date has sought an upper-limit boundary condition for these involuntary processes in the domain of selective attention. Is goal-based selective attention the upper-limit boundary condition of the RIT? Addressing this possibility is informative because selective attention is associated with both executive function/cognitive control (Egner, 2017) and with consciousness.[10]

[8] Ironic effects arise when one thinks about a certain thing, such as a memory or some form of mental imagery, while attempting to not think about that thing. Of import to the present project, according to Wegner (1994), these detected mental contents enter consciousness automatically and involuntarily. (For reviews of ironic processing and thought suppression, see Rassin, 2005; Wegner, 1989.)

[9] With respect to the boundary effects involving emotion, it is obvious that one cannot, by sheer will and without some difficulty, make oneself frightened or ecstatic, as is well known to "method" actors. Hence, subjects might find it easier to follow the instruction "Do not make yourself feel ecstatic [or some other emotional/incentive state]" than the instruction "Do not think of the name of this object" (Cho, Zarolia, Velasquez, & Morsella, 2015).

[10] Much has been theorized regarding the relationship between attention and consciousness (see differing views about this relationship in Koch & Tsuchiya, 2007, and Cohen, Cavanagh, Chun, & Nakayama, 2012). Morsella et al. (2016b) propose that the nature of this relationship depends in large part on one's definition of attention. Today, there are more than a handful of definitions of attention (Tsotsos, 2011). For Oberauer and Hein (2012), there is a low-level

Selective attention based on goals has also been linked to voluntary processing ('the will'), conscious processing, and the sense of the self (Graziano, 2013; James, 1890). With this in mind, one could conclude that perhaps *RIT effects cannot arise for operations that require such a form of attention.* Identifying such a boundary condition would be informative, for the boundary conditions of the RIT effect reveal some of the limits of involuntary processing and thereby shed light on the contributions of conscious processing.

6. The present approach

As noted above, no RITs have focused on attentional processing (specifically, selective attention). Hence, we investigated whether involuntary effects involving selective attention can stem from sets (e.g., to attend to the left) or from stimulus properties (e.g., saliency or processing fluency). In Experiment 1, subjects were presented with a fixation cross surrounded by a clock of 12 words in the stead of numbers. Based on the instructions of the flanker task (Eriksen & Eriksen, 1974; Eriksen & Schultz, 1979), subjects were instructed to focus on the center of the screen and to not count the number of letters of a word at a certain location. (In the original flanker task, subjects were instructed to "respond only to the letter in [a] location and to ignore any and all other letters" [Eriksen & Eriksen, 1974, p. 144].) In this experiment, involuntary counting served as evidence that selective attention shifted involuntarily.

Experiment 2 introduces an RIT (employing Chinese ideographs) that examines the effect of stimulus fidelity and processing fluency. Native English speakers and a separate group of subjects who could read Chinese ideographs were presented with a stimulus array similar to that of Experiment 1. The subjects were instructed to focus on the center of the screen and not read any of the words. Some words were easy to read (e.g., regular Chinese words and English words), and some words were more difficult to read (e.g., Chinese "loan" words and English pseudowords [both defined below]). Would involuntary reading occur more often for fluent words (regular ideographs and English real words) than for disfluent words (Chinese loan words and English pseudowords)? Such a contrast would serve as evidence that selective attention shifted involuntarily as a function of the fluency of the stimulus. It is important to note that this property of the stimulus is supra-perceptual and different in nature from, say, the brightness or loudness of a stimulus.

Our aim was to assess whether attentional phenomena such as selective attention can be influenced involuntarily and systematically through external control. We sought to obtain substantive evidence that, under controlled laboratory conditions designed to minimize artifacts and measurement error, these effects on attentional processing can occur involuntarily and at a reliable rate. This would provide evidence that attentional shifts of this nature can occur involuntarily. For both experiments, we predicted that there would be RIT effects that are nontrivial, reliable, significantly different from zero, and substantive. If the effect fails to arise in this manner, as has occurred for several RITs, then this could illuminate a new boundary condition of the RIT effect. Knowledge of a new boundary condition of the RIT would shed light on the limitations of involuntary processes and also on the contributions of conscious processing. Knowledge of these limitations is important for many subfields of the study of mind and brain, including consciousness, attention, and imagery. We should add that the RIT is the kind of paradigm that, because it builds incrementally on robust phenomena and previous research, has been encouraged by leading researchers in the field (e.g., Fiedler, 2017; Nosek, Spies, & Motyl, 2012).

(*footnote continued*)
form of attention, having certain properties, and a separate, higher-level form of attention, having other properties. One could argue that one of these forms of attention, but not the other forms of attention, is somehow necessary for basic consciousness. Regardless, one must explain how such a form of attention is essential for basic conscious contents (e.g., nausea or smelling a gas leak).

7. Experiment 1

7.1. Method

7.1.1. Subjects

Nineteen San Francisco State University students (M_{age} = 22.67, SD_{age} = 3.92; females = 12) participated for course credit. All subjects were 18 years of age or older. The involvement of human subjects in this study was approved by the Institutional Review Board at San Francisco State University. Prior to participation in the study, all subjects provided written and verbal consent.

7.1.2. Stimuli and apparatus

Stimuli were presented on an Apple iMac computer with a screen measuring 50.8 cm. PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993) was used for the presentation of stimuli and the collection of data. Subjects were seated at a viewing distance of approximately 48 cm. All instructions and prompts were presented in black 36-point Helvetica font on a white background.

All stimuli (n = 240) were composed of a circular array of words surrounding a fixation cross (similar to a clock) in white 36-point Helvetica font on a black background (Fig. 1). Each stimulus "clock" (n = 60) contained one target word and 11 filler words. The target word was always located at either the top position of the clock or the bottom position on the clock. The filler words at the remaining cardinal positions of the clock were matched in frequency ($SUBTLEX_{US}$ Word Frequency Database; Brysbaert & New, 2009) as closely as possible. The remaining filler words were chosen to have the lowest possible word frequency relative to the frequency of the target. The target word was never the highest frequency word in the stimulus array. For a given subject, no word—target or filler—ever appeared more than once.

The target words in the critical trials, words which were composed of three or five (n = 20) letters and which were used successfully in previous research (Merrick et al., 2015), were chosen because they fall below or near the limit of the subitizing range (i.e., three to five letters). Trials with noncritical target words (n = 20), words which were composed of two, four, or six letters, were included to diminish the likelihood that a subject would infer the hypothesis of the study. All filler words, composed of three or five letters, and noncritical target words, were obtained through the $SUBTLEX_{US}$ Word Frequency Database (Brysbaert & New, 2009). To counterbalance whether a particular target word appeared at the top location or bottom location, and whether it was flanked by three-letter or five-letter filler words, four

Fig. 1. Sample stimulus from Experiment 1. Not drawn to scale.

sets of 60 clocks (one for each target word) were created. Each subject was presented with only one of the four sets. The order in which stimuli were presented to each subject was random.

7.1.3. Procedures

Each subject was run individually. Each subject participated in the Top or Bottom condition and was presented with a set of stimuli with the target words in the corresponding top position or bottom position (i.e., the "critical" location). Subjects were informed that they would be presented with a series of images containing words in a circle, similar to a clock, preceded by a fixation cross (+) at the center of the screen. The fixation cross remained at the same position during the presentation of the stimulus array. Subjects were instructed to focus their gaze on the fixation cross and to keep their gaze at that location when each stimulus array appeared. Subjects were then instructed that, for every stimulus array that was presented, they should try not to think of the number of letters in the word at the critical location. Importantly, each subject was told to not attend, for the entire session, to only one location. To avoid confusion, subjects were never told to not attend to more than one location. The experimenter verified that each subject understood the instructions of the task.

Subjects were instructed to press the spacebar if they happened to think of the number of letters in the word at the critical location. Subjects were informed that a beep would be heard if the spacebar was pressed. Subjects were given an opportunity to hear an example of the beep sound when presented with these instructions. Subjects were instructed to keep their hand rested on the spacebar for the duration of the experiment, and were informed that, during a trial, the stimulus array would remain on the screen for a fixed amount of time, whether they pressed the spacebar or not. They were then presented with an example stimulus, on which the critical location was circled in red (Fig. 2). For this example, square shapes were used in place of words so that subjects did not perform any letter counting before the critical trials.

Subjects were asked to repeat the instructions back to the researcher and were given an opportunity to ask questions before proceeding to the critical trials. At the start of each trial (n = 60), subjects were presented with one of the following instructions, depending on the experimental condition: (1) Do NOT think of the number of letters in the word at the TOP, or (2) Do NOT think of the number of letters in the word at the BOTTOM. Subjects pressed the spacebar to advance the trial and be presented with the stimulus (Fig. 3). A fixation cross was presented on the screen for 700 ms, followed by a stimulus (6 s), during which time subjects indicated by pressing the spacebar if they happened to think of the number of letters of the word in the critical location. If the subject pressed the spacebar, a beep was heard. The purpose of the beep was to alert the experimenter, who observed the subject's progress from a short distance and recorded the target word that was presented on each trial. If the subject did not think of the number of letters of the word in the critical location, then they did nothing.

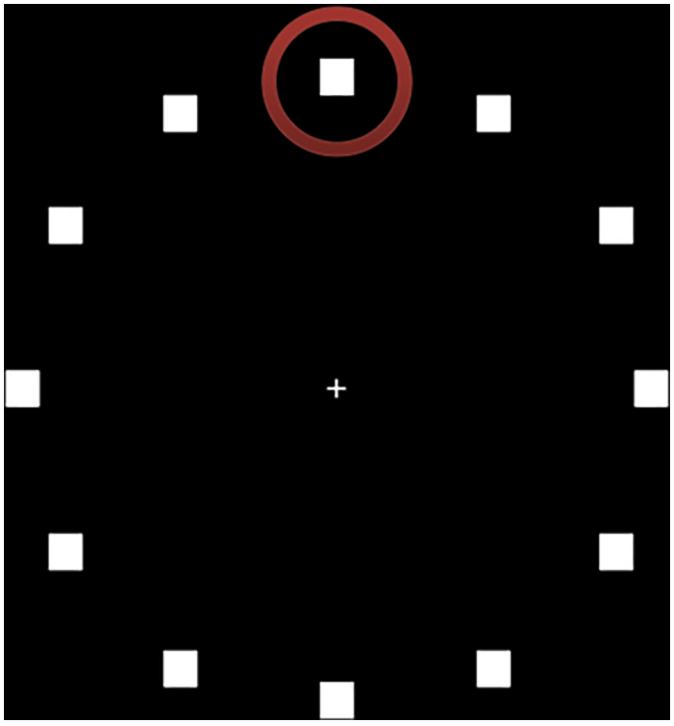

Fig. 2. Training stimulus for Experiment 1, with the target location circled. Not drawn to scale.

After the presentation of each stimulus array, subjects were asked to input by keyboard the number of letters they thought of, or to input "x" if they did not think of the number of letters in the critical word. Subjects were then asked to say the word aloud, or to say "None" if they did not think of the number of letters of the critical word. The researcher manually recorded whether the word spoken by the subject matched the target word for that trial. Subjects uttered the word, and did not input the word by keyboard, to diminish the likelihood of letter counting during the reporting of the word. Last, subjects were instructed to press the "y" key, signifying "yes," if the thought of the number of letters came into mind immediately, and to press the "n" key, signifying "no," if it did not. They pressed the "x" key if they did not think of the number of letters.

After completing the experiment, subjects answered a series of funneled debriefing questions (following the procedures of Bargh & Chartrand, 2000) to assess whether they employed any strategy during the experiment, if they had knowledge of the purpose of the study, or if anything interfered with their performance on the task. These questions were used to determine whether the data from a subject should be omitted from analysis. No data were excluded based on the answers to these questions, but the data from three subjects were excluded from analysis because the subjects failed to follow instructions. For these three subjects, it was obvious to the experimenter that, for some reason, they were not sufficiently engaged, early in the session, with the reading of the instructions or with the learning of when they should or should not press the spacebar. Hence, it was clear to the experimenter that the button presses of these subjects could not be used as an index of the occurrence of an involuntary mental process. For example, one of these subjects believed that the goal of the task was to count willfully the number of letters of words at various locations.

In previous RITs, subjects reported during funneled debriefing that they (a) intended to follow the instructions and (b) attempted some strategies to try to thwart the RIT effect (Allen et al., 2013; Bhangal et al., 2015; Bhangal et al., 2016; Bhangal et al., 2018; Cho et al., 2014; Cushing et al., 2017; Dou et al., 2018; Merrick et al., 2015). Accordingly, for the present RIT, in response to the question in the funneled debriefing, "On each trial, did you feel that you tried (intended) to follow the instructions?", all but one subject indicated something to the effect of "yes." In response to the question, "Did you have a strategy and/or goal in completing this experiment?", 16 subjects conveyed that they adopted a strategy that, they believed, would allow them to not be susceptible to the RIT effect. For example, one subject reported, "I tried to remain focused on the center cross." In this response, "the center cross" refers to the fixation cross. To this same question (that is, "Did you have a strategy and/or goal in completing this experiment?"), two subjects responded "no."

7.1.4. Results

Consistent with our primary prediction, involuntary counting of the critical stimulus occurred on a substantive proportion of the 40 non-filler trials (0.39, SD = 0.22, SE = 0.06, Range = 0.05 to 0.72), a

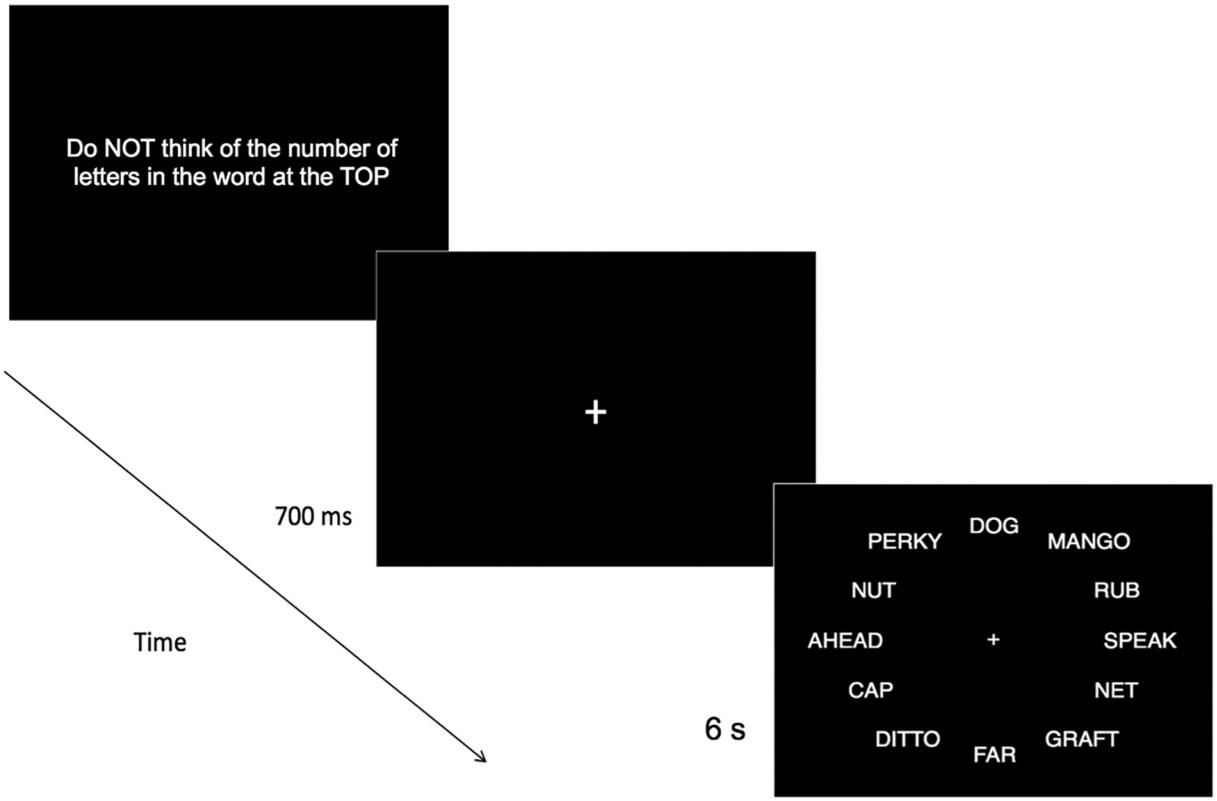

Fig. 3. Schematic depiction of a typical trial in Experiment 1. Not drawn to scale.

proportion that was significantly different from zero, t (15) = 6.93, $p < .001$. Consistent with our secondary prediction, for the Subitizing Range condition, involuntary counting occurred on a proportion of 0.58 of the 20 trials (SD = 0.33, SE = 0.08, Range = 0.05 to 1). This proportion was significantly different from zero, t (15) = 7.11, $p < .001$. For the Outside Range condition, involuntary counting occurred on a proportion of 0.20 of the 20 trials (SD = 0.18, SE = 0.05, Range = 0 to 0.60). This proportion, too, was significantly different from zero, t (15) = 4.46, $p < .001$. Of import, the RIT effect (proportion of trials) in the Subitizing Range condition was significantly different from that in the Outside Range condition, F (1, 15) = 28.13, $p < .0001$ (η_p^2 = 0.65).

An analysis of subjects' trial-by-trial measure of immediacy revealed that, when the involuntary effect arose, it was perceived to be immediate on a mean proportion of 0.81 of the trials (SD = 0.20, SE = 0.05, Range = 0.27 to 1). The mean latency of the involuntary counting was 2568.43 ms (SD = 785.37, SE = 296.84). The mean accuracy of the involuntary counting was high (0.93, SD = 0.15, SE = 0.04, Range = 0.43 to 1).

8. Experiment 2

As mentioned above, in RITs in which subjects are instructed to not think of the name of the to-be-presented object, involuntary subvocalizations are influenced by the word frequency of the name of the object: High-frequency words are more likely to yield an RIT effect than low-frequency words, and the latency of the effect is shorter for the former (Bhangal et al., 2015). This occurred in studies (Bhangal et al., 2015) resembling that of Allen et al. (2013), in which, on a given trial, only one stimulus was presented at a time. This frequency effect was also obtained in the data set of Reyes, Yankulova, Yoo, and Morsella (2017), a study in which two RIT stimuli were presented simultaneously. In this RIT variant, subjects were instructed to not think of the name of *any* of the two objects. The stimulus list included many of the stimuli of Bhangal et al. (2015). There were trials in which one of the objects had a high-frequency name, and the other object had a low-frequency name. Replicating Bhangal et al. (2015), for such trials, the RIT effect was more likely to arise for the high-frequency group of stimuli ($M_{proportion}$ = 0.81, SD = 0.22, SE = 0.03) than for the low-frequency group of stimuli ($M_{proportion}$ = 0.73, SD = 0.22, SE = 0.03), t (46) = 5.88, $p < .0001$.

To build on these previous studies, we investigated the effect on involuntary selective attention from stimulus fidelity and processing fluency. Native English speakers and a separate group of subjects who could read Chinese ideographs were presented with a stimulus array similar to that of Experiment 1. As in Experiment 1, the subjects were instructed to focus on the center of the screen and not read any of the words. Some words were easy to read (e.g., regular Chinese words and English words), and some words were more difficult to read, such as Chinese "loan" words and English pseudowords. Chinese loan words are a case of transliteration in which Chinese characters are combined in order to yield, for example, an English word. For instance, to yield the name "Lucy" through this form of transliteration, the symbol "露" (lu), meaning water droplet, and "西" (xi), meaning East, would be combined, as in "露西". A pseudoword is "a letter string designed to resemble an actual word, even though it is not. Examples include 'BLAR', 'PLOME' and 'TUKE'" (Reisberg, 2015, p. 609).

We assessed whether involuntary reading occurred more often for fluent words (regular ideographs and English real words) than for disfluent words (Chinese loan words and English pseudowords), which would be a case of an involuntary shift in selective attention caused by a nonperceptual feature of the stimuli.

8.1. Method

8.1.1. Subjects

San Francisco State University students (n = 50; Chinese speakers = 25; English speakers = 25; female = 34; M_{age} = 21.91,

SD_{age} = 3.71) participated for course credit or $10. The Chinese speakers were capable of reading and writing Chinese, and the English speakers were not capable of reading and writing Chinese. The Institutional Review Board at San Francisco State University approved the involvement of human subjects in our project. Prior to participation in the study, all subjects provided written and verbal consent.

8.1.2. Stimuli and apparatus

Stimuli were presented on an Apple iMac computer monitor (50.8 cm) with a viewing distance of approximately 48 cm. PsyScope software (Cohen et al., 1993) was used to present the stimuli and record the data. Instructions and questions were written in black 36-point Helvetica font. All stimuli and texts were displayed in black on a white background. There were four sets of stimuli. One set consisted of 30 English real words (Appendix A), from various word classes except prepositions or interjections (e.g., "away"; Brysbaert & New, 2009). A second set (Appendix A) consisted of 30 English pseudowords (e.g., FLUP) and was taken from Rastle, Harrington, and Coltheart (2002); as cited in Jantz, Tomory, Gazzaley, & Morsella, 2013). All English stimuli were presented in black 56-point Helvetica font (approximately 3 cm × 1.5 cm). The English pseudoword stimuli were used successfully in previous research (Jantz et al., 2013).

The other two sets (Appendix A) consisted of 30 ideograph (Chinese) real words from various word classes except prepositions or interjections (e.g., 天气 = "weather"; Cai & Brysbaert, 2010) and 30 ideograph (Chinese) loan words that refer to names in English (e.g., 埃米 = "Amy"; Cai & Brysbaert, 2010). We purposefully avoided ideograph loan words that refer to famous English names, such as the name of a president or celebrity. All ideograph stimuli were presented in black 56-point Song font (approximately 3 cm × 1.5 cm), with each character occupying approximately 1.5 cm × 1.5 cm. The visual angle for both English words and Ideograph words was 1.79° × 3.58° (1.5 cm × 3 cm). The array of both words was presented on the screen with a visual angle of 13.07° x 3.58° (11 cm × 3 cm).

The two types of words within each language were matched on some characteristics. All Chinese stimuli contained only two characters and two syllables. All English real words contained four letters and a maximum of two syllables, and the English pseudowords contained four letters and one syllable. The font size of all stimuli was such that the Chinese words occupied the same amount of space on the computer screen as did the English words. During each trial (Fig. 4), two words were equidistant (3.7 cm) from a fixation cross (+) in the center of the screen. On an English trial, one word from the English real word list was randomly selected along with one word from the English pseudoword list. On an Ideograph trial, one word from the Chinese real-word list was randomly selected to appear with another word from the Chinese loan-word list. For each stimulus array, one word was presented in the upper area, which is 3.7 cm above the fixation cross, or in the lower area, which is 3.7 cm below the fixation cross. An illustration of the upper and lower area was shown to each subject. Placing the stimuli on the upper and lower area of the fixation cross diminished artifacts from spatial incompatibility. Whether the loan word/pseudoword and real word appeared on the upper area or the lower area was randomized.

8.1.3. Procedures

The English and Chinese trials were manipulated within-subjects, with all 120 stimuli intermixed and presented in random order. For example, on the first trial, a subject may see two English stimuli, and, on the second trial, the same subject may see Chinese stimuli. On each trial having English stimuli, subjects never saw two English real words or two English pseudowords. Similarly, on each trial presenting Chinese stimuli, subjects never saw two Chinese real words or two Chinese loan words. Subjects were instructed that they would be presented with two words on each trial, one word above the fixation cross and the other word below the fixation cross. They were further instructed to not read any of the words that were presented, with the instruction "*Do Not Read Any of the Words*" appearing before each trial. See schematic depiction of a typical trial in Fig. 4.

Since subvocalization is a common product of reading, we specifically chose to use the word "Read" instead of "Subvocalize," to reduce confusion and demand characteristics. The "read" here refers to whether or not one extracts phonology and meaning from seeing the words. If the subject did read any of the words that were presented, then the subject indicated this by button press, each time that he or she read a word. For this purpose, there were two buttons provided to the subject: One button corresponded to the word presented in the Upper Area, and one button corresponded to the word presented in the Lower Area. The pressing of these buttons allowed us to quantify the occurrence of involuntary reading. There could be multiple button presses in the same trial, depending on how many times the subject read any of the words. Subjects were informed that they could indicate which word they happened to read by pressing the "z" key or the "/" key. The pairing of the key with the location of the words (i.e., Upper Area or Lower Area) was counterbalanced across subjects. Therefore, the "z" key would indicate the upper word for one subject, and would indicate the lower word for another subject. Both keys were covered with colored paper that had an upper case "U" or "L" written on it, to minimize confusion. The colored papers were pink and yellow. The pairing of colored papers to keys was counterbalanced across subjects. The duration of the presentation of the words (6 s) was the same as that used by Cho, Dou, Reyes, Geisler, and Morsella (2018), an RIT experiment which also presented two stimuli on each trial. Subjects were instructed that, if they did not read either of the words during the trial, then they should not press any buttons. Subjects were also told that the trial would not go by any more quickly or slowly depending on their performance, and that, during the entirety of the trial, they should focus on the fixation cross in the center of the screen.

After each trial, the subject was presented with the question, "*During the trial, did you find yourself paying more attention to one of the areas of the screen? If so, please indicate which area by pressing: Upper (pink/yellow), Lower (yellow/pink) or NEITHER (0 or Zero)*". To reply to the question, subjects pressed the button that corresponded to the location of the word, or inputted the "0" key to indicate that they either paid equal attention to both areas on the screen or did not pay attention to any area. The purpose of the question was to assess if the subject was aware of the occurrence of an attentional shift during the trial, because subjects could involuntarily read both stimuli during the trial. Before the critical trials, subjects completed two practice trials that presented both the English and Chinese condition. The English stimuli used in the practice trials (either ZEAF or MARF paired with either HAVE or MAKE; see Appendix A) were the same for all subjects and never used for the critical trials. The Chinese stimuli used in the practice trials were always, for all subjects, "天气 = Weather" paired with "埃米 = Amy" (Appendix A). These stimuli were never used for the critical trials.

Once subjects completed the experiment, they answered a series of funneled debriefing questions. This questionnaire was designed to help determine whether the data from any subjects should be excluded from analysis. The funneled debriefing questionnaire (following the procedures of Bargh & Chartrand, 2000) included general questions to assess whether subjects (a) were aware of the purpose of the study, (b) were aware of what this experiment was trying to study, (c) had a strategy and/or goal in completing this experiment, (d) had anything interfere with their performance on the task, (e) tried to follow the instructions, (f) knew the meaning of the Ideograph (Chinese) words. The data from one subject were excluded from analysis because the subject had to leave the laboratory during the middle of the session.

It seems that subjects intended to follow the instructions and attempted to thwart the RIT effect. In response to the question in the funneled debriefing, "On each trial, did you feel that you tried (intended) to follow instructions?", all but three subjects indicated something to the effect of "yes." The responses of the three subjects who did not report something to the effect of "yes" were "no," in two cases,

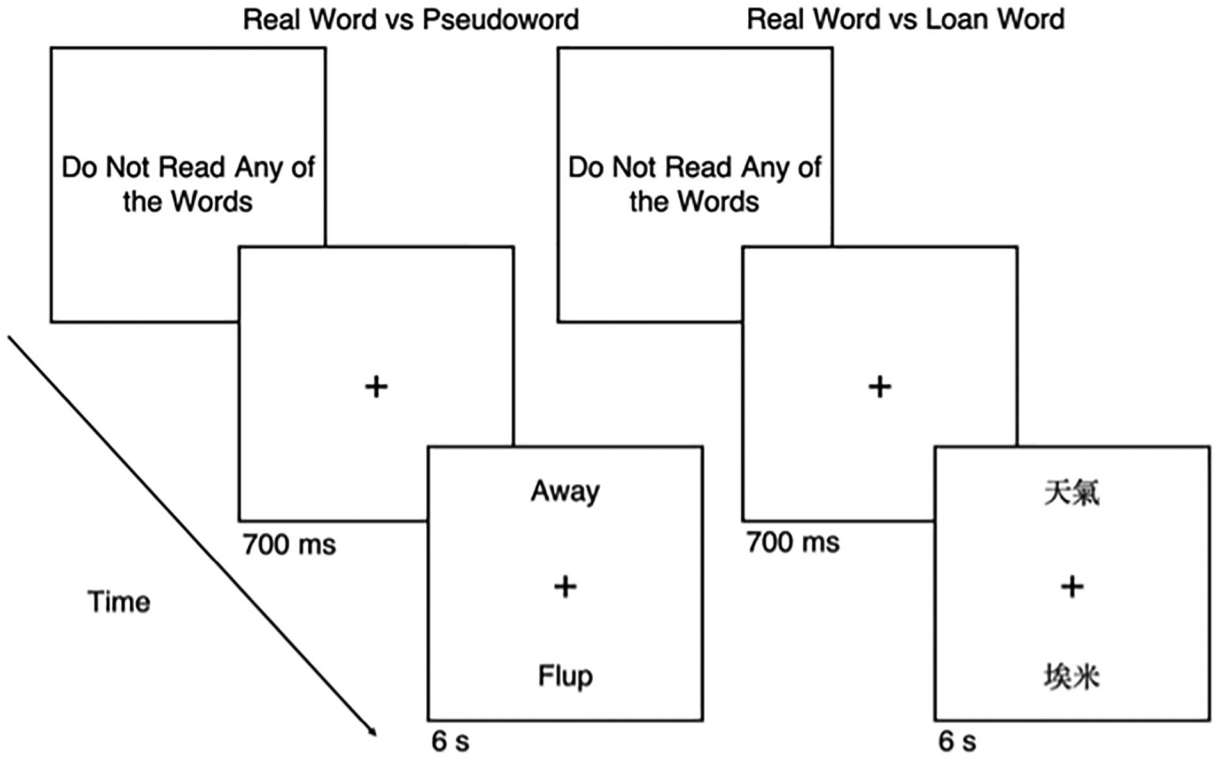

Fig. 4. Schematic depiction of a typical trial in Experiment 2. The stimulus 天气 means "weather," and the stimulus 埃米 means "Amy." Not drawn to scale.

and no response in the third case. In response to the question, "Did you have a strategy and/or goal in completing this experiment?", all but 10 subjects conveyed that they adopted a strategy that, they believed, would allow them to not experience an RIT effect. For example, one subject reported, "My strategy was to focus on the cross [the fixation cross] and not think of anything."

8.1.4. Results

The mean response latency of the button presses, which of course is not an exact measure of the timing of the occurrence of the involuntary reading, was 2546.03 ms (SD = 659.25, SE = 96.16, Range = 1260.55 to 3821.75 ms). For all the conditions displayed in Fig. 5, the mean rate of button presses across trials was significantly different from zero, $ts > 9.14$, $ps < 0.0001$.

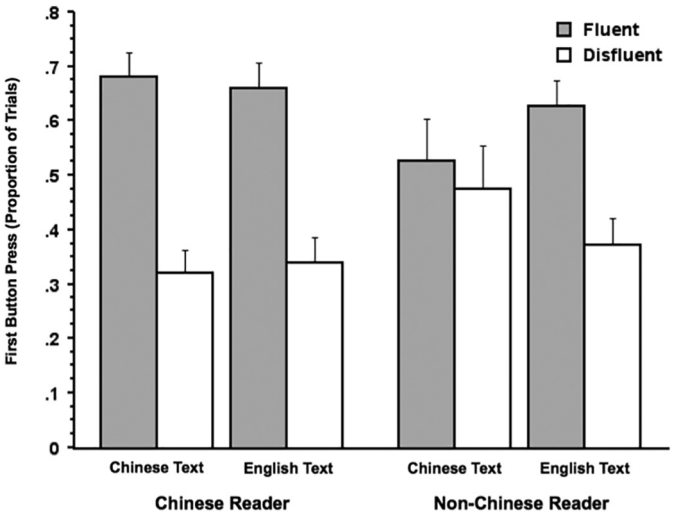

Fig. 5. First button press, indicating involuntary reading, as a function of Reader (Chinese or Non-Chinese), Text Language (Chinese text or English text), and Fluency (Fluent [Chinese regular word or English regular word] and Disfluent [Chinese loan word or English pseudoword]).

We conducted an omnibus ANOVA which had Reader (Chinese or Non-Chinese) as a between-subjects factor, Text Language (Chinese text or English text) as a within-subjects factor, and Fluency (Fluent [Chinese regular word or English regular word] and Disfluent [Chinese loan word or English pseudoword]) as a within-subjects factor. As illustrated in Fig. 5, there was no main effect of Reader, F (1, 37) = 0.51, p = .48, a main effect of Text Language, F (1, 37) = 18.73, p = .0001, and, of most import, a main effect of Fluency, F (1, 37) = 14.23, p = .0006. There were no noteworthy interactions between the factors, $ps > 0.24$, except for a trend between Reader and Fluency, F (1, 37) = 3.26, p = .08. Of main importance, for Chinese readers, there were significantly more button presses, across trials, for the regular words than for the loan words, t (24) = 4.15, p = .0004. Similarly, for the Non-Chinese readers, there were significantly more button presses, across trials, for the regular words than for the pseudowords, t (21) = 2.69, p = .014. English readers did not display such a preference for the Chinese ideographs (i.e., regular words versus loan words), t (13) = 0.32, p = .75. Of note, when confronted with the English words, the Chinese readers, too, had more responses, across trials, to the regular words than to the pseudowords, t (24) = 3.42, p = .002. This is most likely because these subjects, who are university students, are fluent readers of English.

Our critical dependent measure was subjects' responses to the post-trial question ("*During the trial, did you find yourself paying more attention to one of the areas of the screen? If so, please indicate which area by pressing: Upper [pink/yellow], Lower [yellow/pink]) or NEITHER* [*0* or *Zero*]"). The results resembled that of the analysis involving the button presses (Fig. 6), except that, in this analysis, (1) the main effect of Reader was significant, F (1, 48) = 13.73, p = .0005 (η_p^2 = 0.22), (2) the interaction between Reader and Text Language was significant, F (1, 48) = 46.96, $p < .0001$ (η_p^2 = 0.49), and (3) the interaction between Reader and Fluency was significant, F (1, 48) = 5.72, p = .02 (η_p^2 = 0.11). In addition, in this analysis, the interaction between the three factors Reader, Language, and Fluency was significant, F (1,

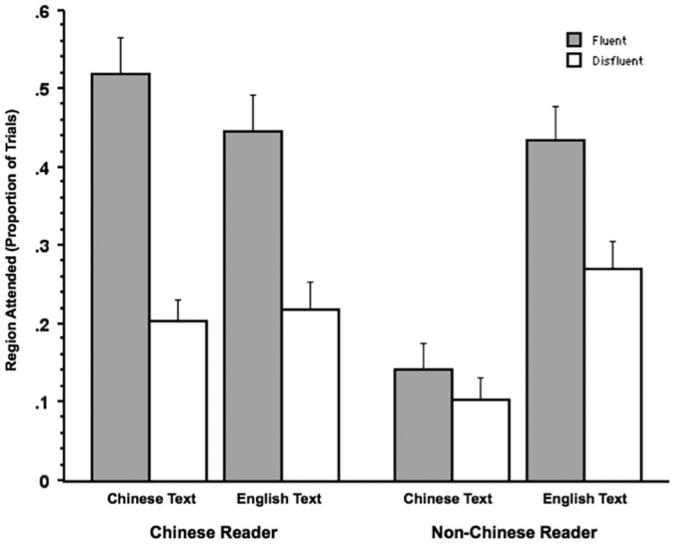

Fig. 6. Region attended (proportion of trials) as a function of Reader (Chinese or Non-Chinese), Text Language (Chinese text or English text), and Fluency (Fluent [Chinese regular word or English regular word] and Disfluent [Chinese loan word or English pseudoword]).

48) = 6.02, p = .02.

Of main importance, for Chinese readers, there was significantly more attending, across trials, to the region associated with regular words than to the region associated with loan words, t (24) = 4.97, p < .0001. For the Non-Chinese readers, there was significantly more attending, across trials, to the regular words than to the pseudowords, t (24) = 2.40, p = .02. English readers did not display such a difference when confronted with Chinese ideographs (i.e., regular words versus loan words), t (24) = 1.77, p = .09. Of note, when confronted with the English words, the Chinese readers, too, attended more, across trials, to the regular words than to the pseudowords, t (24) = 3.29, p = .003. Again, this is most likely because these subjects, who are university students, are fluent readers of English. The response to this post-trial question corresponded to the button-pressing responses on a mean proportion of 0.76 of the trials (SD = 0.22).

9. Discussion

Previous versions of the RIT have revealed how thoughts can be elicited involuntarily through the combination of the activation of sets and the presentation of external stimuli. Here, the data from two experiments reveal that processes involving attention, too, can be controlled in this involuntary manner. In Experiment 1, subjects were instructed to not pay attention to, and not count the number of letters in, a word presented in a critical location (the "critical word"). Throughout all of the scores of trials, the critical location was always the same, to minimize inter-trial carryover effects, confusion on the part of the subject, and other artifacts. Despite the intentions of the subjects, who were instructed to focus their gaze on the fixation cross presented at the center of the screen, involuntary letter-counting of the critical word occurred on a substantive proportion of the trials (39%). Of import, this effect requires an involuntary shift in attention, one that is based on the activation of set, a phenomenon that has been construed as being "top-down" (Miller, 2000). This phenomenon is different from, for example, an involuntary attentional shift to the onset of a salient stimulus (e.g., a bright flash), as occurs in the visual grasp reflex (Sumner & Husain, 2008).

In Experiment 2, involuntary shifts in attention stemmed from the properties of the stimuli: Word stimuli that were high in fluency received involuntary attention more often than did words that were low in fluency. The high-fluency words were real words in English and in Chinese; the low-fluency words were English pseudowords and Chinese loan words. This effect, involving fluency, is different in nature from an attentional effect stemming from lower-level features of the stimulus, for example, the case in which a word presented in red font captures attention when that word is presented along with words presented in standard black font. One could conclude that the present data suggest that the boundary conditions of the RIT effect do not lie in mental operations requiring selective attention.

One limitation of the present experiments is that the dependent measure involves the technique of self-report. Subjects' self-reports could be inaccurate because of a plethora of reasons (see discussion in Morsella, Wilson, et al., 2009), including response bias, demand characteristics, or incorrect introspections (e.g., memory distortions; Block, 2007). However, the accuracy (~90%) of the involuntary counting in Experiment 1 suggests that subjects were in fact reporting their experience of involuntarily attending to, and somehow processing, the stimulus. In line with this conclusion, as mentioned in the Introduction, evidence from other RITs corroborates that subjects' self-reports tend to be accurate and reliable. Future versions of the two present RITs could be combined with neuroimaging technologies to provide additional, neural-based measures (e.g., activation of the neural correlates of reading or of letter-counting) that could corroborate still further the self-reports made by subjects.

The RIT effect supports theoretical views proposing that, in cognition, one is conscious only of the outputs of mental processes but (often) not of the processes themselves. This notion recurred often in the history of psychology and was espoused by Karl Lashley (1956), George Miller (1962), Neal Miller (1959), Helmholtz (1856/1925), and Nisbett and Wilson (1977). Today, some untraditional views regarding the nature of attention are consonant with this notion, as we will discuss next.

Many theorists construe attention as a cause, something that influences information processing in a certain way. For example, the allotment of attention to a perceptual stimulus will cause for the perceptual processing of that stimulus to be enhanced. Other theorists propose that attention should be construed, not as a cause, but rather as an effect (Krauzlis, Bollimunta, Arcizet, & Wang, 2014).

According to Krauzlis et al. (2014), perceptual representations that are of high priority (e.g., associated with bodily needs or activated sets) are perceptually enhanced (involuntarily) compared to the other representations activated at that time. From this standpoint, this relative enhancement is experienced by the observer as a voluntary allocation of attention. Thus, in the traditional view, attention is the cause of perceptual enhancement, but in the view of Krauzlis et al. (2014), attention emerges from the relative perceptual enhancement that high-priority representations receive, a process stemming from unconscious mechanisms in the basal ganglia. Prominent figures such as James (1890), Neisser (1967), and Hochberg (1978), too, had subscribed to an effect-based view of attention, which has the important advantage of avoiding the homunculus problem (see Review in Johnston and Dark, 1986).

The findings of Experiment 1, involving involuntary, set-based attentional shifts are consistent with the view of Krauzlis et al. (2014), because the shift in attention, normally a "top-down" process associated with volition, occurred against the intentions of the subject. From the standpoint of Krauzlis et al. (2014), one could speculate that, in Experiment 1, the set enhanced the critical location and this then led to a deeper processing (letter counting) of the stimulus presented there. Subjects would, according to Krauzlis et al. (2014), experience that they attended more to the critical word. Similarly, from this untraditional standpoint, the fluent stimuli in Experiment 2 were somehow "enhanced" compared to the disfluent stimuli, and this enhancement led to subjects experiencing that they attended more to them.

A future project could, with a single experiment with a within-subjects design, juxtapose the strength of the manipulations from Experiment 1 and Experiment 2. Another future project could, by building on the findings of Experiment 2, investigate whether stimulus

properties other than fluency (e.g., the "pop-out" effect; Treisman & Gelade, 1980) determine which stimulus from a set of stimuli is associated with involuntary entry into consciousness. Experiment 1 involved what could be construed as a form of spatial priming. Would similar RIT effects arise from other forms of pre-trial priming (e.g., semantic priming)? For example, in an RIT in which two stimulus objects (e.g., line drawings of CAT and HAMMER) are presented on each trial, would priming of a concept (e.g., ANIMAL) before the trial influence which of the two objects is associated with involuntary entry (e.g., yielding /k/, /œ/, and /t/)?

In everyday life, it is obvious that most entry into consciousness occurs involuntarily. The RIT was designed to investigate the nature and limitations of such entry. Again, the present data suggest that the boundary conditions of the RIT effect do not lie in mental operations requiring an executive process such as selective attention. The present findings and the theoretical views that they support have implications for various subfields of psychology, including attention, perception, psycholinguistics, cognitive control, and psychopathology. Regarding psychopathology, it is known that, as revealed in the dot-probe task (MacLeod, Mathews, & Tata, 1986), undesired attentional processing (e.g., in anxiety, rumination, obsessions, and addictions) can be triggered by sets and the nature of external stimuli, including the valence of these stimuli (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Hezel & Simpson, 2019; Meyer, 1966; Van Rooijen, Ploeger, & Kret, 2017). The effective control of attention, which is a form of the more general process of cognitive control (Egner, 2017), is an essential component of successful overall self-regulation (Egner, 2017; Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008). By understanding the stimulus conditions and set-related conditions under which attentional control fails, thereby causing involuntary cognitions to enter consciousness, one can develop a more complete theory of the interaction between voluntary and involuntary processing in the process of self-regulation.

Credit authorship contribution statement

Katelyn Gardner: Conceptualization, Methodology, Supervision, Writing - original draft. **Erica B. Walker:** Conceptualization, Methodology, Supervision, Writing - original draft. **Yanming Li:** Conceptualization, Methodology, Supervision, Writing - original draft. **Adam Gazzaley:** Conceptualization, Methodology, Supervision, Writing - original draft. **Ezequiel Morsella:** Conceptualization, Methodology, Supervision, Writing - original draft.

Acknowledgment

Lara Krisst conducted the pilot study in which subliminal stimuli (orthographs) were presented as the stimuli in the Reflexive Imagery Task.

Appendix A. Stimulus list

English real word	English non-word	Ideograph real word	English translation	Ideograph loan word	English translation
Away	Brun	我们	We	安吉	Angie
Body	Crin	确实	Indeed	巴里	Barry
Come	Darf	分析	Analysis	邦尼	Bonnie
Ever	Flup	自己	Ourself	查德	Chad
Fine	Gelp	抱歉	Sorry	科迪	Cody
Game	Gerp	离开	Leave	伊恩	Ian
Help	Hame	接受	Accept	艾琳	Irene
Inch	Irms	孩子	Kid	唐娜	Donna
Joke	Jong	当然	Forsure	摩根	Morgen
Knee	Jort	可以	Can	伊桑	Ethan
Lift	Kels	安全	Safe	吉纳	Gina
List	Lerg	认为	Consider	戈登	Gordon
Mind	Lirm	而且	And	霍利	Holly
Name	Maff	勇气	Courage	基恩	Keith
Nose	Marp	明白	Understand	肯特	Kent
Only	Nint	学习	Learn	拉里	Larry
Park	Nirm	现在	Now	玛吉	Maggie
Past	Phiv	听说	Heard	尼尔	Neil
User	Rull	水平	Horizontal	欧文	Owen
Ring	Rurn	奶酪	Cheese	佩里	Perry
Step	Shov	微笑	Smile	昆西	Quincy
Vase	Terg	朋友	Friend	帕克	Parker
Time	Tunk	告诉	Tell	萨莉	Sally
Town	Veam	房子	House	谢莉	Shelly
Very	Warl	出来	Come out	鲁斯	Ruth
Warm	Yarp	电梯	Elevator	泰勒	Tyler
What	Yash	释放	Release	文斯	Vince
Yarn	Zean	晚餐	Dinner	佐伊	Zoe
Zone	Zere	偶像	Idol	芬恩	Finn
Does	Flis	早安	Morning	吉恩	Jin

References

Ach, N. (1905/1951). Determining tendencies: Awareness. In D. Rapaport (Ed.). *Organization and pathology of thought* (pp. 15–38). New York: Columbia University Press (Original work published in 1905.).

Allen, A. K., Krisst, L., Montemayor, C., & Morsella, E. (2016). Entry of involuntary conscious contents from ambiguous images. *Psychology of Consciousness: Theory, Research, and Practice, 3*, 326–337.

Allen, A. K., Wilkins, K., Gazzaley, A., & Morsella, E. (2013). Conscious thoughts from reflex-like processes: A new experimental paradigm for consciousness research. *Consciousness and Cognition, 22*, 1318–1331.

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis, & C. M. Judd (Eds.). *Handbook of research methods in social and personality psychology* (pp. 253–285). Cambridge, England: Cambridge University Press.

Bargh, J. A., & Morsella, E. (2008). The unconscious mind. *Perspectives on Psychological Science, 3*, 73–79.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*, 1–24.

Bhangal, S., Cho, H., Geisler, M. W., & Morsella, E. (2016). The prospective nature of

voluntary action: Insights from the reflexive imagery task. *Review of General Psychology, 20*, 101–117.
Bhangal, S., Merrick, C., Cho, H., & Morsella, E. (2018). Involuntary entry into consciousness from the activation of sets: Object counting and color naming. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2018.01017.
Bhangal, S., Merrick, C., & Morsella, E. (2015). Ironic effects as reflexive responses: Evidence from word frequency effects on involuntary subvocalizations. *Acta Psychologica, 159*, 33–40.
Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences, 30*, 481–548.
Brysbaert, M., & New, B. (2009). Moving beyond Ku_era and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.
Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One, 5*(6), e10729.
Cho, H., Dou, W., Reyes, Z., Geisler, M. W., & Morsella, E. (2018). The reflexive imagery task: An experimental paradigm for neuroimaging. *AIMS Neuroscience, 5*, 97–115.
Cho, H., Godwin, C. A., Geisler, M. W., & Morsella, E. (2014). Internally generated conscious contents: Interactions between sustained mental imagery and involuntary subvocalizations. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01445.
Cho, H., Zarolia, P., Gazzaley, A., & Morsella, E. (2016). Involuntary symbol manipulation (Pig Latin) from external control: Implications for thought suppression. *Acta Psychologica, 166*, 37–41.
Cho, H., Zarolia, P., Velasquez, A., & Morsella, E. (2015). Cognitive- versus emotion-based involuntary cognitions: An informative contrast for the reflexive imagery task. *Poster presented at the annual convention of the Association for Psychological Science, New York*.
Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences, 16*, 411–417.
Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, & Computers, 25*, 257–271.
Cushing, D., Gazzaley, A., & Morsella, E. (2017). Externally controlled involuntary cognitions and their relations with other representations in consciousness. *Consciousness and Cognition, 55*, 1–10.
Desender, K., van Opstal, F. V., & van den Bussche, E. (2014). Feeling the conflict: The crucial role of conflict experience in adaptation. *Psychological Science, 25*, 675–683.
Dostoevsky, F. (1863/2008). *Winter notes on summer impressions.* London: One-world Classics.
Dou, W., Li, Y., Geisler, M. W., & Morsella, E. (2018). Involuntary polymodal imagery involving olfaction, audition, touch, taste, and vision. *Consciousness and Cognition, 62*, 9–20.
Egner, T. (2017). *The Wiley handbook of cognitive control.* West Sussex, UK: John Wiley and Sons.
Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.
Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics, 25*, 249–263.
Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science, 12*, 46–61.
Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences, 39*, 1–77 Target Article.
Garcia, A. C., Bhangal, S., Velasquez, A. G., Geisler, M. W., & Morsella, E. (2016). Metacognition of working memory performance: Trial-by-trial subjective effects from a new paradigm. *Frontiers in Psychology, 7*, 927. https://doi.org/10.3389/fpsyg.2016.00927.
Gazzaley, A., & D'Esposito, M. (2007). Unifying prefrontal cortex function: Executive control, neural networks and top-down modulation. In B. L. Miller, & J. L. Cummings (Eds.). *The human frontal lobes: Functions and disorders* (pp. 187–206). New York: Guilford Press.
Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54*, 493–503.
Goodhew, S. C. (2017). What have we learned from two decades of object-substitution masking? Time to update: Object individuation prevails over substitution. *Journal of Experimental Psychology: Human Perception & Performance, 43*(6), 1249–1262.
Graziano, M. S. A. (2013). *Consciousness and the social brain.* New York: Oxford University Press.
Helmholtz, H. (1856/1925). Treatise of physiological optics: Concerning the perceptions in general. In T. Shipley (Ed.). *Classics in psychology* (pp. 79–127). New York: Philosophy Library.
Hezel, D. M., & Simpson, H. B. (2019). Exposure and response prevention for obsessive-compulsive disorder: A review and new directions. *Indian Journal of Psychiatry, 61*, S85–S92.
Hochberg, J. E. (1978). *Perception.* Englewood Cliffs, New Jersey: Prentice-Hall.
James, W. (1890). *The principles of psychology.* New York, NY: Dover.
Jantz, T. K., Tomory, J. J., Gazzaley, A., & Morsella, E. (2013). Subjective aspects of action control for delayed actions: Action-related imagery. *Journal of Mental Imagery, 37*, 21–48.
Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology, 37*, 43–75.
Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences, 11*, 16–22.
Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences, 18*, 457–464.
Lashley, K. S. (1956). Cerebral organization and behavior. *Proceedings of the Association for Research in Nervouse and Mental Diseases, 36*, 1–18.
Lewin, K. (1935). *A dynamic theory of personality.* New York: McGraw-Hill.
Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65*, 272–292.
MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*, 15–20.
Mason, M. F., Norton, M. I., van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science, 315* (393-345).
McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin, 136*, 188–207.
Merrick, C., Farnia, M., Jantz, T. K., Gazzaley, A., & Morsella, E. (2015). External control of the stream of consciousness: Stimulus-based effects on involuntary thought sequences. *Consciousness and Cognition, 33*, 217–225.
Meyer, V. (1966). Modification of expectations in cases with obsessional rituals. *Behavior Research and Therapy, 4*, 273–280.
Miller, N. E. (1959). Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Vol. Ed.), *Psychology: A study of a science. Vol. 2. Psychology: A study of a science* (pp. 196–292). New York: McGraw-Hill.
Miller, G. A. (1962). *Psychology: The science of mental life.* New York: Adams, Bannister, & Cox.
Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience, 1*, 59–65.
Miller, B. L., & Cummings, J. L. (2007). *The human frontal lobes: Functions and disorders* (2nd ed.). New York: Guilford Press.
Mitchell, J. P., Heatherton, T. F., Kelley, W. M., Wyland, C. L., Wegner, D. M., & Macrae, C. N. (2007). Separating sustained from transient aspects of cognitive control during thought suppression. *Psychological Science, 18*, 292–297.
Morsella, E. (2005). The function of phenomenal states: Supramodular interaction theory. *Psychological Review, 112*, 1000–1021.
Morsella, E., Godwin, C. A., Jantz, T. J., Krieger, S. C., & Gazzaley, A. (2016a). Homing in on consciousness in the nervous system: An action-based synthesis. *Behavioral and Brain Sciences, 39*, 1–17 Target Article.
Morsella, E., Godwin, C. A., Jantz, T. K., Krieger, S. C., & Gazzaley, A. (2016b). Passive frame theory: A new synthesis. *Behavioral and Brain Sciences, 39*, 44–70.
Morsella, E., Gray, J. R., Krieger, S. C., & Bargh, J. A. (2009). The essence of conscious conflict: Subjective effects of sustaining incompatible intentions. *Emotion, 9*, 717–728.
Morsella, E., Wilson, L. E., Berger, C. C., Honhongva, M., Gazzaley, A., & Bargh, J. A. (2009). Subjective aspects of cognitive control at different stages of processing. *Attention, Perception, & Psychophysics, 71*, 1807–1824.
Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.
Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science, 3*, 400–424.
Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II: Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*, 615–631.
Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science, 21*, 164–169.
Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., ... Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology, 10*(1), e1001251.
Questienne, L., Atas, A., Burle, B., & Gevers, W. (2018). Objectifying the subjective: Building blocks of metacognitive experiences in conflict tasks. *Journal of Experimental Psychology: General, 147*, 125–131.
Rassin, E. (2005). *Thought suppression.* Amsterdam, The Netherlands: Elsevier.
Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology, 55A*, 1339–1362.
Reisberg, D. (2015). *Cognition: Exploring the science of the mind* (6th ed.). New York: Norton.
Reyes, Z., Yankulova, J. K., Yoo, S. H., & Morsella, E. (2017). Resilience and involuntary processing of valenced stimuli: The factor of approach/avoidance orientation. *Poster presented at the annual convention of the Association for Psychological Science, Boston*.
Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vols. 1 and 2.* Cambridge, MA: Massachusetts Institute of Technology.
Schacter, D. L., & Tulving, E. (1994). *Memory systems.* Cambridge, MA: The MIT Press.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
Sumner, P., & Husain, M. (2008). At the edge of consciousness: Automatic motor activation and voluntary control. *The Neuroscientist, 14*, 474–486.
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.
Tsotsos, J. K. (2011). *A computational perspective on visual attention.* Cambridge, MA: MIT Press.
Uznadze, D. (1966). *The psychology of set.* New York: Consultants Bureau.
Van Rooijen, R., Ploeger, A., & Kret, M. E. (2017). The dot-probe task to measure emotional attention: A suitable measure in comparative studies? *Psychonomic Bulletin & Review, 24*, 1686–1717.
Wegner, D. M. (1989). *White bears and other unwanted thoughts.* New York: Viking/Penguin.
Wegner, D. M. (1994). Ironic processes of thought control. *Psychological Review, 101*, 34–52.
Wyland, C. L., Kelley, W. M., Macrae, C. N., Gordon, H. L., & Heatherton, T. F. (2003). Neural correlates of thought suppression. *Neuropsychologia, 41*, 1863–1867.