The Impact of Stimuli Color in Lexical Decision and Semantic Word Categorization Tasks

Margarida V. Garrido[a], Marília Prada[a], Cláudia Simão[b] and Gün R. Semin[c,d]

[a]Instituto Universitário de Lisboa (ISCTE-IUL), CIS – IUL

[b]Católica Lisbon School of Business and Economics, Universidade Católica Portuguesa

[c]William James Center for Research, ISPA-Instituto Universitário

[d]Faculty of Social and Behavioral Sciences, Utrecht University

Note: Margarida V. Garrido, and Marília Prada, Department of Social and Organizational Psychology, ISCTE-IUL, CIS-IUL, Lisboa, Portugal; Cláudia Simão, Católica Lisbon School of Business and Economics, Universidade Católica Portuguesa; Gün R. Semin, William James Center for Research, ISPA-Instituto Universitário, Lisboa Portugal and Faculty of Social and Behavioral Sciences, Utrecht University.

* The research reported here was supported by the Fundação para a Ciência e Tecnologia, Portugal, with grants awarded to the first (PTDC/MHC-PCN/5217/2014), second (LISBOA-01-0145-FEDER-

028008) and the fourth (IF/00085/2013/CP1186/CT0001) authors.

The authors would like to thank Filipa Tavares for her help in data collection.

Address for correspondence:

M. V. Garrido, ISCTE-IUL, Av. das Forças Armadas, Office AA109, 1649-026, Lisboa, Portugal.

E-mail: margarida.garrido@iscte-iul.pt

Abstract

In two experiments, we examined the impact of color on cognitive performance by asking participants to categorize stimuli presented in three different colors: red, green, and grey (baseline). Participants were either asked to categorize the meaning of words as related to the concepts of "go" or "stop" (Experiment 1) or to indicate if a neutral verbal stimulus was a word or not (lexical decision task, Experiment 2). Overall, we observed performance facilitation in response to go stimuli presented in green (vs. red or grey) and performance inhibition in response to go stimuli presented in red. The opposite pattern was observed for stop-related stimuli. Importantly, results also indicated that color might also be used to categorize neutral stimuli. Overall, these findings provide support to the green-go and red-stop color associations and test the potential functional autonomy acquired by these colors and the boundary conditions to their effects on stimuli categorization.

Keywords: colors; green; red; stop; go; word categorization tasks.

Color perception may seem to be a merely physiological process based on the sensitivity to different wavelengths of light. However, colors may assume complex communicative functions (Elliot & Maier, 2012). Indeed, color may have an important impact on how we feel, think or behave (for reviews, see Elliot, 2015; Elliot & Maier, 2012, 2014). In our everyday lives, we are used to interpret and respond to color cues. Traffic lights are the most evident example: the green light indicates that is safe to proceed, amber warns that the signal is about to change to red, and red prohibits traffic from proceeding.

This color code, usually known as the Red, Amber and Green (RAG) rating, has also been adapted to other labeling systems and applied in different domains, including the assessment of work or academic performance (e.g., Given, Hannigan, & McGrath, 2016), as well as safety hazards (e.g., Ahmad, Hashim, & Hassim, 2016) or health risks (Carrington, Kok, Jansen, & Stewart, 2013; Prideaux, Conway O'Brien, & Chevassut, 2014). Another widespread application is the nutritional traffic light (e.g., Cecchini & Warin, 2016; Trudel, Murray, Kim, & Chen, 2015) in which the color associated to nutrients such as fat, sugar or salt, signals their quantity in the composition of the product (e.g., foods with green indicators are healthier than those with red ones). Instead of the three discrete colors, other labeling schemes include color codes running from green to red. For example, in the EU these codes are used to indicate the energy consumption of appliances (i.e., the energy class is labeled by a color code associated to letters, such that more energy-efficient appliances are associated to green, and less energy-efficient ones are associated to red); and the recommended age for videogames according to content (e.g., those with green labels may be used by children, whereas those with red are only recommended for adults as they can include violent or sexual content).

Moreover, in our daily life we often associate colors to specific meanings (Adams & Osgood, 1973; Aslam, 2006; Hupka, Zaleski, Otto, Reidl, & Tarabrina, 1997) or use color terms metaphorically (e.g., Meier & Robinson, 2005; Redondo & Plaza, 2007; Rouw, Case, Gosavi, & Ramachandran, 2014). For instance, a non-perceptual state such as the experience of an emotion may be conceptualized

perceptually through its association to a color (e.g., anger is associated to "seeing red", Fetterman, Robinson, & Meier, 2012).

The meanings and effects of red and green (and blue)

Our particular interest is in the effects of two opposing colors (Hurvich & Jameson, 1957) – red and green. We provide below a brief overview of the research on the influence of these colors on behavior.

Research using both self-report and implicit measures has shown that the colors red and green are linked to several concepts – aside from anger, red seems to enhance perceived attractiveness, danger and dominance, whereas green has been associated with creativity and safety (for a review, see Mentzel, Schücker, Hagemann, & Strauss, 2017).

Other studies have manipulated the color of the environment in which participants were placed (e.g., Küller, Ballal, Laike, Mikellides, & Tonello, 2006; Kwallek & Lewis, 1990; Stone, 2001), or the color of the stimuli they were asked to evaluate. For example, Maier et al. (2013) showed that a target-person (job applicant) wearing a red (vs. green) shirt or tie was judged as less intelligent and as less likely to be hired. In another study, Moller, Elliot and Maier (2009) used a reaction time paradigm to examine the associations between colors and words that varied not only in valence, but also in content (i.e., failure vs. success, see Experiment 2). Depending on the condition, participants were asked to categorize the words (presented in red, green and white) as success/failure related or as positive/negative. Results revealed positive associations (i.e., facilitation) between red and failure as well as with negative words and negative associations with success and positive words. Importantly, green was only positively associated with success-related words and this association did not extend to valence. In another study using a reaction time paradigm, Pravossoudovitch, Cury, Young and Elliot (2014) examined the associations between red/danger and green/safety by asking participants to categorize both danger (e.g., emergency) and safety-related (e.g., family) words that were presented in red, green or gray (baseline). Overall, participants were faster (and more accurate) to evaluate danger words presented in red (vs. green or gray), and to evaluate safety-

related words presented in green (vs. red or gray). Notably, green did not inhibit the processing of danger-related words and red did not inhibit the processing of safety-related words.

The studies assessing the impact of primary colors - red versus green (or versus blue) on performance (for a review, see Elliot, 2015) have also included indicators of motor (e.g., Elliot & Aarts, 2011; Elliot, Maier, Binser, Friedman, & Pekrun, 2009; Meier, D'Agostino, Elliot, Maier, & Wilkowski, 2012; Payen et al., 2011) and cognitive performance, namely attention (Hochman, Henik, & Kalanthroff, 2018), information processing (Soldat, Sinclair, & Mark, 1997), and memory (for a review, see Dzulkifli & Mustafar, 2013). The pattern of findings obtained was not clear-cut: some studies have found that red (vs. blue or green) inhibits performance, whereas others show that it facilitates it (for a review, see Mehta & Zhu, 2009). Other studies did not observe any differences produced by color in performance (Arthur, Cho, & Muñoz, 2016; Larsson & von Stumm, 2015).

To resolve the observed discrepancy in the results, several factors have been examined. Examples include the role played by individual (e.g., ego depletion - Bertrams, Baumeister, Englert, & Furley, 2015; gender - Gnambs, Appel, & Batinic, 2010) or contextual factors (e.g., romance vs. achievement contexts - Meier et al., 2012); and type or difficulty of the task (for a review, see Xia, Song, Wang, Tan, & Mo, 2016). For example, Mehta and Zhu (2009) suggested that the impact of color on performance would depend on the type of task and found that blue enhanced performance on creative tasks (e.g., generation of creative uses for a brick), whereas red enhanced performance on detailed-oriented tasks (e.g., memory and proofreading tasks). Xia et al. (2016) further examined how task difficulty moderated the impact of color on performance and found that red only enhanced performance on a detail-oriented task when the task was easy. In contrast, blue enhanced performance on a creativity-oriented task, irrespective of task difficulty. In a series of six studies Mehta and Zhu (2009) showed that the impact of color on cognitive task performance is mediated by the activation of alternative motivations, such that blue activated approach and red activated avoidance motivations. For example, in one of the studies the authors observed that the background color of the computer screen influenced performance (response times of correct answers to

anagrams), such that participants were faster to respond to avoidance-related words when the background was red (vs. blue or white) and were faster to respond to approach-related words when the background was blue (vs. red or white). However, a subsequent study with a larger sample using the same procedure and materials, failed to replicate these particular findings (Steele, 2014).

Other studies have examined the associations between red and green as relevant cues for behavioral action such as stop and go (e.g., Bergum & Bergum, 1981; Chan & Courtney, 2001). For example, Ng and Chan (2018) asked a group of designers and non-designers to select the color that best associated with a given concept. There were three concept categories (total 38 concepts): warnings (e.g., danger, safe, toxic), required action (e.g., go, stop, close) and signs and equipment status (on, off, out of order). Overall, the results showed that the color selected more often by both groups of participants was red followed by green, suggesting a higher number of concept associations for these colors. Although participants selected red and green as the colors best associated to stop and go respectively, the red-stop association was stronger.

In a recent study, Hochman et al. (2018) used a modified stop signal task with red and green traffics lights as stopping-related and going-related environmental cues. Results showed that participants were faster to stop when a red light was present in comparison to a green or black (neutral cue) light. In no-stop trials, participants were slower to respond when a red traffic light was present, and no differences were found between green and black traffic light. These differences between green and red were interpreted as evidence for automatic inhibition due to a red-stop association (rather than a facilitation due to a green-go association).

Overview

Given the prominent role of colors as environmental cues, in triggering adaptive behavior, our goal was to further investigate the role of red and green, on cognitive performance. In two experiments, we tested the red/stop and green/go associations by examining the role played by stimuli color (green and red) in word categorization tasks. The main goal of the current work was to replicate previous findings regarding the red/stop and green/go associations, using a different task

and baseline conditions that would potentially permit uncovering the strength of both associations. Moreover, we wanted to examine whether red and green might have acquired functional autonomy (e.g., Allport, 1937), that is, a new emergent function that is independent from the one from which they developed. In other words, if green and red would facilitate or inhibit performance irrespective of the stimulus content.

In Experiment 1, we orthogonally manipulated the color (red and green) and the go or stop related meaning of the stimuli. Additionally, we also used the color grey as a baseline condition, to further examine the facilitatory or inhibitory nature of the color effects. In Experiment 2, we tested the effect of color on performance using a lexical decision task with stimuli that are unrelated to go/stop. The aim was to examine if these associations have become so automated that they can affect performance on a task content that is not associated with red or green on the surface or if they emerge only when the color is directly relevant in the narrowly defined context.

Experiment 1

To further establish the go/green and stop/red associations, we used a simple categorization task asking participants to classify a set of previously piloted go/stop related words that were presented in green and red. We expected to observe an interaction between word type and color, such that participants would be faster and more accurate to respond to go-related words presented in green and to stop-related words in red. Moreover, to better understand the nature of the associations between the concepts and the colors, we included words in grey as a baseline condition. This condition would potentially clarify whether green facilitates the processing of go-related words and red of stop-related words or if, alternatively, red inhibits the processing of go-related words and green the processing of stop related words. Notably, this design also permits to examine whether both facilitation and inhibition processes are operating. Finally, we have included neutral words (i.e., words unrelated to the concepts of stopping or going), presented in the three colors, to provide a preliminary test to the potential functional autonomy acquired by these colors. If

the color effect generalizes regardless of the stimuli content, then we would expect for example a higher frequency of "go responses" when these neutral words were presented in green (vs. grey or red).

1.1 Method

1.1.1 Participants and Design

A sample of 76 university students (86.8% Females, M_{age} = 20.89, *SD* = 3.21) volunteered to participate in a laboratory study for partial course credit. All participants reported being native speakers (L1) and having normal or corrected to normal vision.

The design included two within-participants factors: 3 (Word: stop; neutral; go) x 3 (Color: gray; green; red). All the manipulations, measures used, and data exclusions are reported. Sample size was based on prior work with a similar design (e.g., Pravossoudovitch et al., 2014).

1.1.2 Materials

Materials included 10 Go-related words (e.g., to walk; to move; to advance), 10 Stop-related words (e.g., to cease; to immobilize; to brake) and 10 words unrelated to Go/Stop (e.g., to look; to peel; to pour)[1]. To select these words, we conducted a pilot study (*n* = 32) in which 60 words from a published word database (Freitas & Albuquerque, 2007) were evaluated using a semantic differential (1 = *Stop*; 7 = *Go*). The sub-sets of words differed on this dimension, such that the highest ratings were observed for go (*M* = 5.63, *SE* = 0.20), followed by neutral (*M* = 4.06, *SE* = 0.15), and stop (*M* = 2.64, *SE* = 0.21) words, $F(2,62) = 60.68$, $MSE = 71.46$, $p < .001$, $\eta_p^2 = .662$ (all $p < .001$, Bonferroni adjustment for multiple comparisons). The average number of letters per word (*M* = 7.13, *SE* = .30) did not differ between the three types of words, $F < 1$.

1.1.3 Procedure

Participants arrived in the laboratory to collaborate in a study aiming to "examine how people categorize words". All the procedures were conducted in line with the ethical guidelines of

[1] All materials and data are available at https://osf.io/ekvty/

the host institution. After written consent was obtained, the instructions were presented with E-Prime 2.0 software (Psychology Software Tools, 2012).

Each participant completed two blocks of 90 trials. Participants were informed that words would be presented in the center of the screen and that their task would simply be to indicate if each word was related to the concepts of Go or Stop by pressing the corresponding key (e.g., Z = *Go;* M = *Stop,* response key assignment was counterbalanced). The first task consisted of a training phase that included six trials (three Go words and three Stop words, two of which presented in grey, two in green and two in red). These stimuli were not used in the experimental trials. During the training phase, feedback about response accuracy and response time was provided. Next, instructions stated that although no feedback would be provided in the subsequent tasks, participants should continue to answer as fast and as accurately as possible.

Participants then proceeded to the experimental task. In the first block of 90 trials, each of the 30 words was randomly presented in three colors (i.e., grey, green and red[2]) in ARIAL font (18 points) against a white background. Each trial started with a fixation point ("*", 500 ms), and was followed by the target stimuli (visible until a response was registered). The inter-trial interval was 1000 ms. Upon completion, participants were informed that the next task would be similar, but the response-keys would be reversed (e.g., Z = *Stop;* M = *Go*). The second block was also preceded by a training phase and used the same materials and procedure as the first one. The experiment lasted approximately 15 minutes. At the end of the experiment, participants were thanked for their participation and fully debriefed.

1.2 Results and Discussion

Only correct responses to the target stimuli provided in a 300 to 1500 ms interval were analyzed (e.g., Bargh & Chartrand, 2000). This meant that 3.17 % of the data were removed. None of the participants showed below chance level (i.e., .50) accuracy in the Stop and Go word sets. Next, we present results regarding the impact of color in the categorization of:

[2] The colors used were predefined in E-Prime 2.0 software, specifically "green" (RGB: 0,128,0), "red" (RGB: 255,0,0) and "gray" (RGB: 128,128,128).

(a) Stop and Go words: we analyzed response accuracy and response times using a Linear Mixed Model analysis, with a 2 (type of stimulus) x 3 (color) within-subjects design.

(b) Neutral words: because the task comprised a forced-choice response between two options (i.e., go vs. stop-related), we only analyzed the proportion of neutral words that were categorized as associated with go (for a similar procedure, see Farias, Garrido, & Semin, 2013; Lakens, Semin, & Garrido, 2011) according to the color in which they were presented – Linear Mixed Model analysis with 3 levels (color). We also replicated this analysis with RTs as the dependent variable.

1.2.1 Impact of Color: Stop/Go Words

Response accuracy

To analyze the proportion of correct responses, we conducted a Generalized Linear Mixed Model, entering color and type of stimulus as fixed factors after controlling for the item length (i.e., number of letters of each word). Subjects and items were entered as random effects to adjust for possible variation. Using the method from Nakagawa and Schielzeth (2013) we observed that the inclusion of subjects and item random intercepts improved the model, as the variance explained by both fixed and random factors (Conditional R^2 = .27) was larger than the variance explained only by the fixed factors (Marginal R^2 = .04)[3].

The overall hit proportion was .93 (*SE* = 0.01). Results revealed a main effect of type of stimulus, *b* = -0.75, *SE* = 0.16, *Z* = -4.65, *p* < .001, 95% CI [-1.07; -0.43]. Go words obtained a higher hit proportion (*M* = .93, *SE* = 0.01) than Stop words (*M* = .92, *SE* = 0.01). Additionally, both main effects of color were significant - green vs. red: *b* = -1.16, *SE* = 0.15, *Z* = -7.83, *p* < .001, 95% CI [-1.44;

[3] We repeated the Generalized Linear Mixed Model for the response accuracy analysis, controlling for the effect of block. Color and type of stimulus were entered as fixed factors. Additionally we controlled for the main effect of block and the effect of word length. Subject and item were entered as random effects. Similarly to the main analysis, the main effects of stimulus type (*b* = -0.75, *SE* = 0.16, *Z* = -4.65, *p* < .001, 95% CI [-1.07; -0.43]), green vs. red (*b* = -1.16, *SE* = 0.15, *Z* = -7.83, *p* < .001, 95% CI [-1.44; -0.87]), and gray vs. red (*b* = -0.42, *SE* = 0.16, *Z* = -2.60, *p* < .001, 95% CI [-0.73; -0.10]) were statistically significant. However, the main effect of block was not statistically significant (*b* = -0.03, *SE* = 0.08, *Z* = -0.38, *p* < .705, 95% CI [-0.20; 0.13]). As expected, the interaction effects of a) type of stimulus and green vs. red (*b* = 2.09, *SE* = 0.22, *Z* = 9.53, *p* < .001, 95% CI [1.66; 2.52]) and b) type of stimulus and gray vs. red (gray vs. red: *b* = 0.88, *SE* = 0.22, *Z* = 4.08, *p* < .001, 95% CI [0.46; 1.30]) remained statistically significant.

-0.87]; gray vs. red: *b* = -0.42, *SE* = 0.16, *Z* = -2.60, *p* = .009, 95% CI [-0.73; -0.10] - meaning that participants were overall less accurate in categorizing words presented in green (*M* = .91, *SE* = 0.01) than words presented in red (*M* = .93, *SE* = 0.01) or in gray (*M* = .93, *SE* = 0.01).

As expected, both interactions of a) type of stimulus and green vs. red, *b* = 2.09, *SE* = 0.22, *Z* = 9.53, *p* < .001, 95% CI [1.66; 2.52]; and b) type of stimulus and gray vs. red, *b* = 0.88, *SE* = 0.22, *Z* = 4.08, *p* < .001, 95% CI [0.46; 1.30] were significant (see Figure 1).

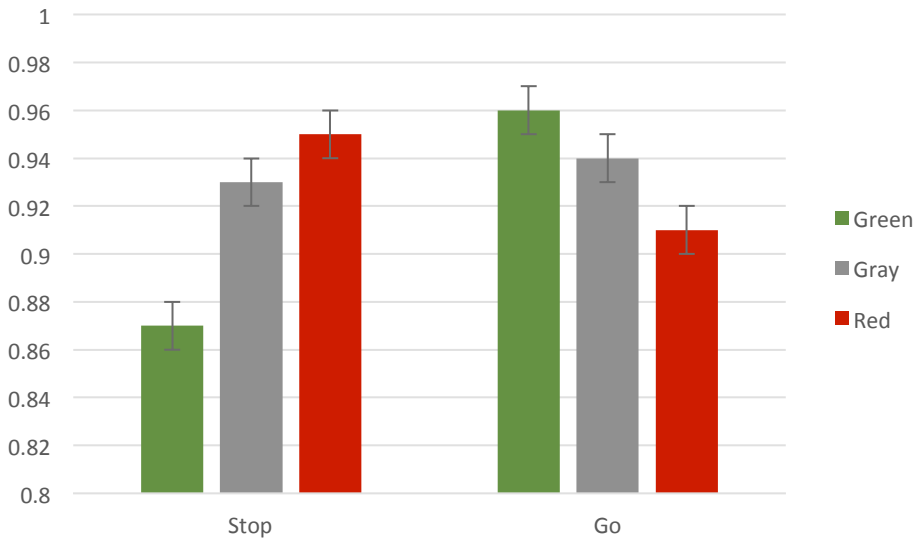

Figure 1. Hit rates (proportion) for Stop and Go words according to presentation color. Error bars represent standard errors.

For stop-related words, hit rates were lower for words in green (*M* = .87, *SE* = 0.01) than in gray (*M* = .93, *SE* = 0.01), *b* = -0.75, *SE* = 0.13, *p* < .001, or red (*M* = .95, *SE* = 0.01), *b* = -1.18, *SE* = 0.15, *p* < .001. The difference between red and gray was also significant, *b* = 0.42, *SE* = 0.16, *p* = .024. In contrast, for go-related words, hit rates were higher for words presented in green (*M* = .96, *SE* = 0.01) than gray (*M* = .94, *SE* = 0.01), *b* = 0.47, *SE* = 0.17, *p* = .017, or red (*M* = .91, *SE* = 0.01), *b* = 0.92, *SE* = 0.16, *p* < .001. The difference between red and gray was also significant, *b* = -0.45, *SE* = 0.14, *p* = .004.

Response times

Response times were entered as the dependent variable in a Linear Mixed Model analysis.

Color and stimulus type were entered as fixed effects, while subject and item were entered as random effects, after controlling for the effect of item length. The combination of both fixed and random effects explained a larger proportion of variance (Conditional R^2 = .27) when compared to the fixed effects alone (Marginal R^2 = .01; Nakagawa & Schielzeth, 2013).[4]

The overall response time was 706 ms (*SE* = 12). The main effect of type of stimulus did not reach significance, *b* = 1.88, *SE* = 7.00, *t* < 1, *p* = .789, 95% CI [-11.84; 15.60], but the main effects of color were significant - green vs. red: *b* = 29.85, *SE* = 6.67, *t*(7066.07) = 4.48, *p* < .001, 95% CI [16.77; 42.92]; gray vs. red: *b* = 14.22, *SE* = 6.55, *t*(7014.87) = 2.17, *p* = .030, 95% CI [1.38; 27.07]. Participants showed faster responses for green colored stimuli (*M* = 699, *SE* = 11), followed by red (*M* = 707, *SE* = 11) and gray (*M* = 711, *SE* = 11) colored stimuli.

As expected, the interaction between type of stimulus and green vs. red (*b* = -75.99, *SE* = 9.37, *t*(7040.99) = -8.11, *p* < .001, 95% CI [-94.36; -57.62]) and the interaction between type of stimulus and gray vs. red (*b* = -20.19, *SE* = 9.32, *t*(7031.97) = -2.17, *p* = .030, 95% CI [-38.45; -1.93]) were both significant (see Figure 2).

[4] We repeated the Generalized Linear Mixed Model for the response time analysis, controlling for the effect of block. Color and type of stimulus as fixed factors, controlling for the main effect of block and the effect of word length. Subject and item were again entered as random effects. The main effect of type of stimulus was not statistically significant, *b* = 1.67, *SE* = 6.97, *t*(5767.25) = 0.24, *p* = .810, 95% CI [-11.99; 15.33]. As expected, both main effects of color (green vs. red: *b* = 29.39, *SE* = 6.62, *t*(7063.33) = 4.44, *p* < .001, 95% CI [16.41; 42.37]; gray vs. red: *b* = 13.95, *SE* = 6.51, *t*(7015.57) = 2.14, *p* = .032, 95% CI [1.20; 26.70]) were statistically significant. Additionally, the main effect of block was also statistically significant (*b* = -37.81, *SE* = 3.79, *t*(7008.16) = -9.99, *p* < .001, 95% CI [-45.24; -30.39]), meaning that participants were faster in the second block (*M* = 687, *SE* = 11) when compared to the first block (*M* = 725, *SE* = 11). The interaction effects of stimulus type and green vs. red (*b* = -75.42, *SE* = 9.31, *t*(7038.35) = -8.10, *p* < .001, 95% CI [-93.65; 57.17]) and of stimulus type and gray vs. red (*b* = -19.70, *SE* = 9.25, *t*(7029.39) = -2.13, *p* = .033, 95% CI [-37.83; -1.57]) remained statistically significant.

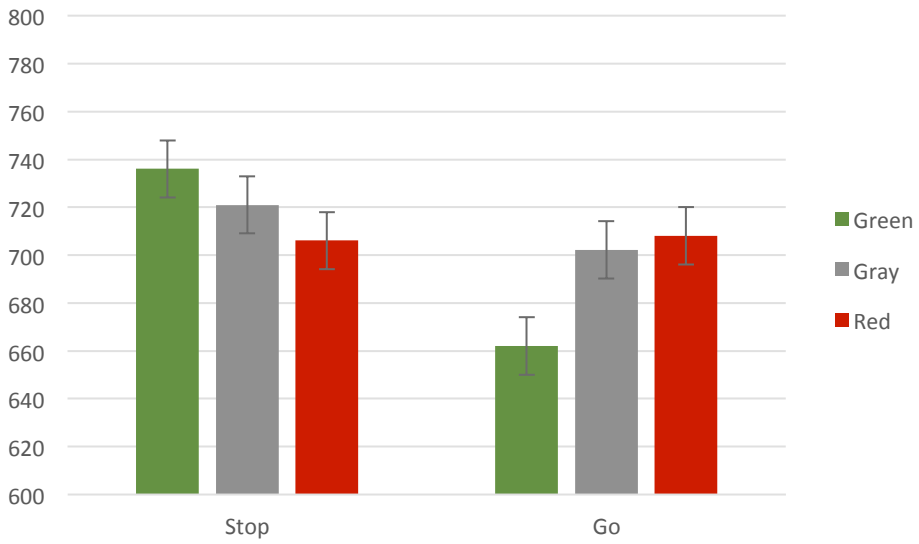

Figure 2. Response times (ms) for Stop and Go words according to presentation color. Error bars represent standard errors.

As shown in Figure 2, participants were slower to categorize stop-related words presented in green (M = 736, SE = 12) than in gray (M = 721, SE = 12), b = 15.87, SE = 6.69, p = .047, or red (M = 706, SE = 12), b = 30.07, SE = 6.65, p < .001. The difference between response to red and grey stop-related words was marginal, b = -14.20, SE = 6.53, p = .076. In contrast, participants were faster to respond to go-related words presented in green (M = 662, SE = 12) than in gray (M = 702, SE = 12), b = -40.18, SE = 6.55, p < .001, or red (M = 708, SE = 12), b = -46.18, SE = 6.60, p < .001. Response times to go-related words presented in red were equivalent to those presented in gray, b = -6.00, SE = 6.64, p = .639. Moreover, pairwise comparisons showed that go-related words presented in grey (M = 702, SE = 12) were categorized faster than stop-related words presented in grey (M = 721, SE = 12, $Mdiff$ = -18.31, SE = 6.98, p = .009, 95% CI [-32.00; -4.62]).

To further examine the role of the specific semantic content of the go/stop words in participants' response times, we conducted a second Linear Mixed Model analysis entering response time as the dependent variable, color as fixed effect, and the ratings of stop/go of each item (obtained in the pilot study, with higher ratings corresponding to go-related categorization) as a continuous fixed predictor. Again, subject and item were entered as random effects to adjust for

possible variation and item length as a control variable. As in the previous model, the combination of both fixed and random effects explained a larger proportion of variance (Conditional R^2 = .27) when compared to the fixed effects alone (Marginal R^2 = .01; Nakagawa & Schielzeth, 2013).

The main effect of green vs. red, *b* = -8.27, *SE* = 4.69, *t*(7043.84) = -1.77, *p* = .078, 95% CI [-17.46; 0.92], and gray vs. red (*b* = 4.15, *SE* = 4.66, *t*(7033.56) = 0.89, *p* = .373, 95% CI [-4.98; 13.29], were marginal and not significant, respectively. Also, the main effect of the ratings of stop/go was not significant, *b* = 0.38, *SE* = 3.49, *t* < 1, *p* = .914, 95% CI [-6.46; 7.22]. However, as expected, the interaction between these ratings and green vs. red, *b* = -36.30, *SE* = 4.67, *t*(7036.15) = -7.77, *p* < .001, 95% CI [-45.46; -27.14], and gray vs. red, *b* = -8.74, *SE* = 4.65, *t*(7020.20) = -1.88, *p* = .060, 95% CI [-17.85; 0.36] were significant and marginal, respectively.

Simple pairwise comparisons showed that participants were slower when categorizing words with lower ratings of stop/go presented in green (*M* = 734, *SE* = 12) than in gray (*M* = 719, *SE* = 12), *b* = 15.13, *SE* = 6.70, *p* = .024, or red (*M* = 706, *SE* = 12), *b* = 28.03, *SE* = 6.66, *p* < .001. The difference between response times to words with lower ratings of stop/go presented in gray and red was also significant, *b* = 12.90, *SE* = 6.55, *p* = .049. In contrast, participants were faster to respond to words with higher ratings of stop/go when presented in green (*M* = 662, *SE* = 12) than in gray (*M* = 702, *SE* = 12), *b* = -39.98, *SE* = 6.52, *p* < .001, or red (*M* = 707, *SE* = 12), *b* = -44.57, *SE* = 6.56, *p* < .001. Response times of words presented in gray were equivalent to those presented in red, *b* = -4.59, *SE* = 6.61, *p* = .487.

1.2.2 Impact of Color on Neutral Words

Based on the number of go responses of each participant we computed the proportion of go responses given to neutral stimuli presented in red, green and grey. Again, subject and item were entered as random effects, after controlling for the effect of item length. The results showed a main effect of green vs. red, *b* = 0.67, *SE* = 0.94, *Z* = 7.12, *p* < .001, 95% CI [0.48; 0.85], and a main effect of gray vs. red, *b* = 0.20, *SE* = 0.94, *Z* = 2.11, *p* = .035, 95% CI [0.01; 0.38]. This means that neutral words presented in green obtained a higher proportion of go responses (*M* = .44, *SE* = 0.02), than those

presented in gray (*M* = .37, *SE* =0.02), *p* < .001, or in red (*M* = .34, *SE* = 0.02), *p* < .001. The proportion of go responses to words presented in red was also lower than in gray, *p* = .089 (pairwise comparisons with Tukey adjustment). Not surprisingly, the categorization of neutral words (803 ms; *SE* = 14) was 97 ms slower than the overall time required to categorize stop and go words (see 1.2.1), suggesting the difficulty of the categorization of neutral words as stop or go-related. Response times to neutral words yielded a significant main effect of green vs. red, *b* = -51.08, *SE* = 12.79, *t*(1421.57) = -3.99, *p* < .001, 95% CI [-76.17; -25.98] and a non-significant main effect of gray vs. red, *b* = -8.36, *SE* = 13.28, *t* < 1, *p* = .529, 95% CI [-34.41; 17.69]. Specifically, stimuli presented in green (*M* = 771, *SE* = 15) was categorized significantly faster than stimuli presented in red (*M* = 822, *SE* = 16), *p* < .001, and gray (*M* = 814, *SE* = 16), *p* = .001. Gray and red did not significantly differ, *p* = .803 (Tukey adjustment).[5]

Overall, these findings suggest as expected, that the color green facilitates the processing of go-related words and inhibits the processing of stop-related words. The opposite pattern was observed for the color red, that is, facilitation of stop-related words and inhibition of go-related words. Moreover, when the words were unrelated to either go or stop (i.e., neutral words) the proportion of go responses was higher and the response times were faster when the words were presented in green suggesting the impact of color even in the absence of the word's relevant semantic content. Experiment 2 aims to further explore the potential functional autonomy acquired by these colors by examining their effect on the categorization of neutral and pseudo-words.

[5] We rerun the analysis controlling for the effect of block. Mean accuracy was higher and response times were faster in the second block. The overall pattern of main effects and interactions remained the same. Based on the number of go responses of each participant we computed the proportion of go responses given to neutral stimuli presented in red, green and grey. Again, subject and item were entered as random effects, after controlling for the main effect of block and the effect of item length. The results showed a main effect of green vs. red, *b* = 0.67, *SE* = 0.94, *Z* = 7.13, *p* < .001, 95% CI [0.49; 0.86], a main effect of gray vs. red, *b* = 0.20, *SE* = 0.94, *Z* = 2.11, *p* = .035, 95% CI [0.01; 0.38], and a main effect of block, *b* = 0.31, *SE* = 0.08, *Z* = 4.07, *p* < .001, 95% CI [0.16; 0.46]. Response times for neutral words yielded a significant main effect of green vs. red, *b* = -50.01, *SE* = 12.49, *t*(1415.81) = -4.01, *p* < .001, 95% CI [-74.48; -25.54] and a non-significant main effect of gray vs. red, *b* = -4.48, *SE* = 12.98, *t* < 1, *p* = .730, 95% CI [-29.89; 20.93]. The main effect of block was statistically significant, *b* = -86.39, *SE* = 10.28, *t*(1447.64) = -8.41, *p* < .001, 95% CI [-106.54; -66.25].

Experiment 2

This experiment included a lexical decision task (LDT) in which neutral words and pseudo words (generated from re-arranging the letters of the original words) were presented in the three colors. The words used were unrelated to the typical concepts (e.g., danger/safety, stop/go) associated to red and green. Hence, this experiment could potentially clarify whether green and red have become functionally autonomous, that is, whether they can facilitate or inhibit performance in the absence of the original stimuli they have been associated with. Two predictions are possible. The first would be that the color effect would have become so automatized that color on its own is generalized across stimuli, namely functionally independent (in the current case, performance facilitation in response to stimuli presented in green and performance inhibition to stimuli presented in red). The second prediction suggests that there may be boundary conditions to the color main effect, and that this will only emerge when it is relevant to the stimuli (i.e., when the stimuli are somehow associated with the color).

2.1 Method

2.1.1 Participants and Design

A sample of 61 university students (85.2% Females, M_{age} = 21.39, *SD* = 3.32) volunteered to participate in a laboratory study for partial course credit. All participants reported being native speakers (L1) and having normal or corrected vision.

The design included two within-participants factors: 2 (Stimuli Type: pseudo-word; word) x 3 (Color: gray; green; red). All the manipulations, the measures used, and data exclusions are reported.

2.1.2 Materials

The word stimuli set included 10 concrete words selected from a published word database based on valence and familiarity ratings (7-point rating scales, Garcia-Marques, 2003). The words selected (e.g., “roldana” and “trincha”, which correspond to “pulley” and “brush” in English) were moderate in valence (*M* = 3.99, SD = 0.20) and familiarity (*M* = 4.06, SD = 0.30). These words served

as the input for the 10 pseudo-words set. Specifically, we re-arranged the letters of each word taking into account the orthographic and phonologic structure of the participants language, so that the pseudo-word could be pronounced (e.g., Domingos & Garcia-Marques, 2008; Farias, Garrido, & Semin, 2016; Godinho & Garrido, 2016). For example, the pseudo-words deriving from "roldana" and "trincha" were "dolnara" and "chintra", respectively. On average, each stimulus contained 7 letters.

2.1.3 Procedure

Participants were invited to the laboratory to collaborate in a study aimed at "exploring how people perceive and evaluate words". All the procedures followed the ethical guidelines of the host institution. After written consent was obtained, the instructions were presented with E-Prime 2.0 software (Psychology Software Tools, 2012).

Each participant completed two blocks of 60 trials (order counterbalanced between participants). Participants were informed that verbal stimuli would be presented in the center of the screen and that their task would simply be to indicate if that stimuli was a Word or a Pseudo-Word by pressing the corresponding key (e.g., Z = *Word;* M = *Pseudo-Word,* response keys assignment in counterbalanced order). As in Experiment 1, the first task was a training phase that included six trials (three words and three pseudo-words, two of which presented in grey, two in green and two in red). These stimuli were not used in the experimental trials. During the training phase, feedback about response accuracy and response time was provided. Next, instructions stated that although no feedback would be provided in the subsequent tasks, participants should continue to answer as fast and as accurately as possible. Participants then proceeded to the first block of 60 trials: each stimulus (10 words and 10 pseudo-words) was randomly presented in three colors (i.e. grey, green and red), in ARIAL font (18 points) against a white background. A typical trial of the lexical decision task started with a fixation point ("*", 500 ms), and was followed by the target stimuli (visible until a response was registered). The inter-trial interval was 1000 ms. Upon completion, participants were informed that the next task would be similar, but the response-keys would be reversed (e.g., Z =

Pseudo-Word; M = *Word*). The second block was also preceded by a training phase and used the same materials and procedure as the first one. The experiment lasted approximately 12 minutes. At the end of the experiment, participants were thanked for their participation and fully debriefed.

2.2 Results and Discussion

Only correct responses to the target stimuli provided in a 300 to 1500 ms interval (e.g., Bargh & Chartrand, 2000) were analyzed. This meant that 3.05 % of the data were removed. Then, we calculated the proportion of overall hits per participant and excluded those (n = 1) with performance below chance level (i.e., .50). The final sample included 60 participants. Results for hits and response times are presented independently in the following sections.

2.2.1 Response Accuracy

To analyze the proportion of correct responses, we conducted a Generalized Linear Mixed Model, entering color and type of stimulus as fixed factors after controlling for item length. Subject and item were entered as random effects to adjust for possible variation. Additionally, the intercepts for subject and item improved the model, as the variance explained by both fixed and random factors (Conditional R^2 = .56) was larger than the variance explained only by the fixed factors (Marginal R^2 = .03; Nakagawa & Schielzeth, 2013).[6]

The overall hit proportion was .87 (*SE* = 0.01). The main effect of type of stimulus, was not significant, *b* = 0.26, *SE* = 0.19, *Z* = 1.37, *p* = .170, 95% CI [-0.11; 0.63], with participants showing similar hit rates for words (*M* = .89, *SE* = 0.01) and pseudo-words (*M* = .85, *SE* = 0.01). The main effects of color - green vs. red: *b* = -0.12, *SE* = 0.14, *Z* = -0.89, *p* = .373, 95% CI [-0.39; 0.15]; gray vs.

[6] A Generalized Linear Mixed Model was conducted with proportion of correct responses as the dependent variable, and both color and stimulus type as fixed effect. We entered also the main effect of block and controlled for word length. The random intercepts for subject and item were taken into account in the analysis to adjust for possible variation. Results revealed that the main effects of type of stimulus (*b* = 0.26, *SE* (0.19), *Z* = 1.38, *p* = .169, 95% CI [-0.11; 0.64]) or of color (green vs. red: *b* = -0.12, *SE* (0.14), *Z* = -0.89, *p* = .372, 95% CI [-0.39; 0.15]; gray vs. red: *b* = -0.18, *SE* (0.14), *Z* = -1.30, *p* = .194, 95% CI [-0.45; 0.09]) were not significant. However the main effect of block was statistically significant (*b* = 0.35, *SE* (0.09), *Z* = 4.11, *p* < .001, 95% CI [0.18; 0.52]). Participants showed a higher proportion of hits in the second block (*M* = .88, *SE* = 0.01) when compared to the first block (*M* = .86, *SE* = 0.01). As expected from the main analysis, bot interaction effects were statistically significant (stimulus type X green vs. red: *b* = 0.71, *SE* (0.21), *Z* = 3.34, *p* < .001, 95% CI [0.29; 1.12]; stimulus type X gray vs. red: *b* = 0.66, *SE* (0.21), *Z* = 3.16, *p* = .002, 95% CI [0.25; 1.07]).

red: *b* = -0.18, *SE* =0.14, *Z* = -1.30, *p* = .196, 95% CI [-0.44; 0.09] - were not significant.

Importantly, results also showed a significant interaction between the type of stimulus and the green vs. red colors, *b* = 0.70, *SE* = 0.21, *Z* = 3.33, *p* < .001, 95% CI [0.29; 1.12], and a significant interaction between the type of stimulus and the gray vs. red colors, *b* = 0.66, *SE* = 0.21, *Z* = 3.15, *p* = .002, 95% CI [0.25; 1.07] (see Figure 3).

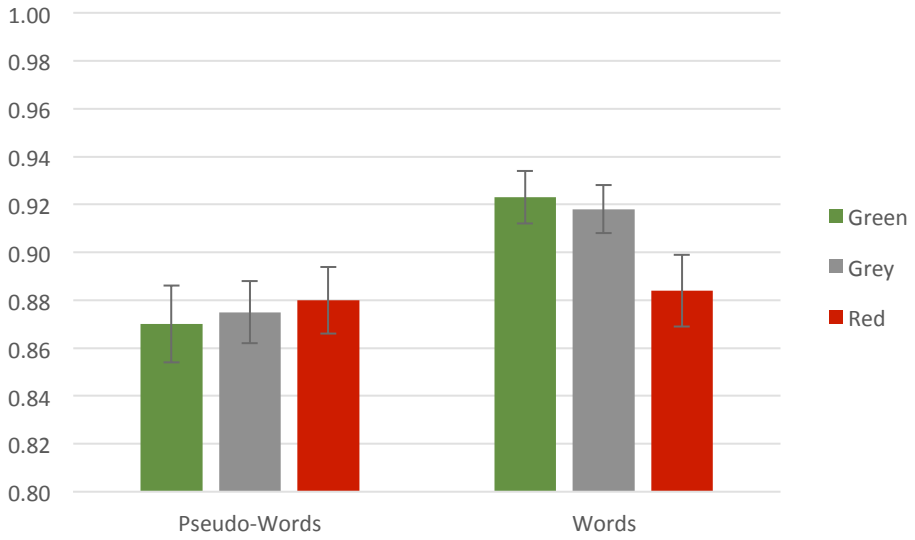

Figure 3. Hit rates (proportion) according to type (words vs. pseudo-words) and color (gray, green, red) of target stimuli. Error bars represent standard errors.

Specifically, hit rates in response to pseudo-words did not vary according to the color of the stimuli: green (*M* = .85, *SE* = 0.02) vs. gray (*M* = .84, *SE* = 0.02), *b* = 0.05, *SE* = 0.13, *p* = .918; and green vs. red (*M* = .86, *SE* = 0.02), *b* = -0.12, *SE* = 0.13, *p* = .659, and gray vs. red, *b* = -0.17, *SE* = 0.13, *p* = .416. In contrast, hit rates were higher for words presented in green (*M* = .91, *SE* = 0.02) and gray (*M* = .90, *SE* = 0.02) in comparison to red (*M* = .87, *SE* = 0.02), green vs. red, *b* = 0.63, *SE* = 0.17, *p* < .001; gray vs. red, *b* = 0.52, *SE* = 0.17, *p* = .005. The difference between green and gray was not significant, *b* = 0.11, *SE* = 0.18, *p* = .811.

2.2.2 Response Times

Response times were entered as the dependent variable in a Linear Mixed Model analysis. Color and stimulus type were entered as fixed effects, and subject and item as random effects, after

controlling for the effect of word length. The combination of both fixed and random effects explained a larger proportion of variance (Conditional R^2 = .36) when compared to the fixed effects alone (Marginal R^2 = .05; Nakagawa & Schielzeth, 2013).[7]

The overall response time in the task was 699 ms (*SE* = 14). Results indicated a main effect of type of stimulus, with faster responses when the target was a word (*M* = 656, *SE* = 13) versus a pseudo-word (*M* = 743, *SE* = 13), *b* = -63.37, *SE* = 7.72, *t*(4116.04) = -8.21, *p* < .001, 95% CI [-78.50; -48.25]. The main effect of green vs. red was significant, *b* = 26.05, *SE* = 7.17, *t*(5253.67) = 3.64, *p* < .001, 95% CI [12.01; 40.09], showing that participants were overall faster when stimuli were presented in green (*M* = 698, *SE* = 13) than in red (*M* = 700, *SE* = 13). The main effect of gray vs. red was not significant, *b* = 7.35, *SE* = 7.17, *t*(5231.16) = 1.03, *p* = .306, 95% CI [-6.71; 21.40].

The interaction between type of stimulus and green vs. red colors was also significant, *b* = -56.62, *SE* = 10.03, *t*(5223.84) = -5.65, *p* < .001, 95% CI [-76.26; -36.97], but the interaction between type of stimulus and gray vs. red was not significant, *b* = -14.25, *SE* = 10.04, *t*(5216.47) = -1.42, *p* = .156, 95% CI [-33.92; 5.42] (see Figure 4).

[7] Data from response times was analyzed with a Linear Mixed Model where color and type of stimulus were entered as fixed factors, controlling for the main effect of block and the effect of word length. Subject and item were again entered as random effect. The main effect of type of stimulus was statistically significant, *b* = -64.52, *SE* = 7.63, *t*(3987.13) = -8.46, *p* < .001, 95% CI [-79.47; -49.57]. The main effect of green vs. red was also statistically significant, *b* = 25.92, *SE* = 7.00, *t*(5246.65) = 3.70, *p* < .001, 95% CI [12.20; 39.64] but the main effect of gray vs. red was not, *b* = 7.22, *SE* = 7.01, *t*(5224.75) = 1.03, *p* = .302, 95% CI [-6.51; 20.95]). Additionally, the main effect of block was also statistically significant (*b* = -62.55, *SE* = 4.00, *t*(5228.07) = -15.64, *p* < .001, 95% CI [-70.39; -54.71]), meaning that participants were faster in the second block (*M* = 669, *SE* = 13) when compared to the first block (*M* = 731, *SE* = 13). Similarly to the main analysis, the interaction effect of stimulus type and green vs. red (*b* = -56.63, *SE* = 9.79, *t*(5217.60) = -5.78, *p* < .001, 95% CI [-75.81; -37.43]) was statistically significant but the interaction of stimulus type and gray vs. red (*b* = -14.57, *SE* = 9.81, *t*(5210.48) = -1.49, *p* = .138, 95% CI [-33.78; 4.65]) was not.

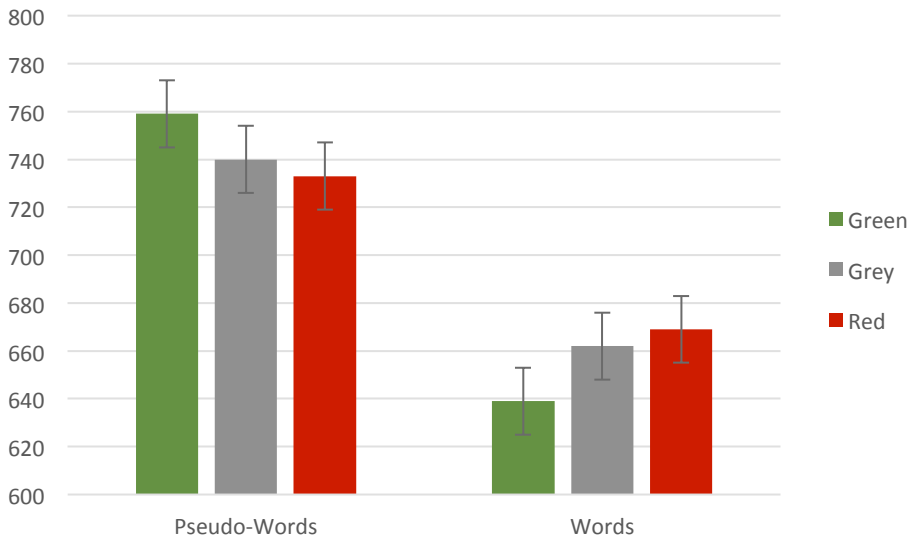

Figure 4. Response Times (ms) according to type (words vs. pseudo-words) and color (gray, green, red) of target stimuli. Error bars represent standard errors.

Slower response times were observed when pseudo-words were presented in green (M = 758, SE = 14) than in grey (M = 739, SE = 14), b = 17.92, SE = 7.57, p = .047, or red (M = 732, SE = 14), b = 25.91, SE = 7.53, p = .002. Responses to pseudo-words presented in gray versus in red did not differ, b = 7.99, SE = 7.54, p = .539. The opposite pattern was found for words: faster responses when they were presented in green (M = 638, SE = 14) than in grey (M = 661, SE = 14), b = -23.41, SE = 6.58, p = .001, or red (M = 668, SE = 14), b = -30.77, SE = 6.64, p < .001. Again, no differences were observed for responses to words presented in gray versus in red, b = -7.37, SE = 6.65, p = .510.

General Discussion

Contextual factors have been shown to influence human behavior and cognition. In this work we examined how one of these factors – color – namely red and green impacts on cognitive performance, and test the potential functional autonomy acquired by these colors as well as the boundary conditions to their effects. Previous research has shown that a given color may be associated to various meanings (e.g., red is associated to “danger” and to “stop”; green is associated to “growth” and to “go”). Still, the activation of a given meaning may be more likely depending on

the presence of other colors in the context - for instance, perceiving red along with green is likely to activate the stop and go meanings. These associations are not surprising given the common use of colors codes used in traffic light systems and have already been documented in self-report studies (e.g., Ng & Chan, 2018).

In two experiments, we examined the influence of the color of different verbal stimuli in the performance of categorization and lexical decision tasks. In Experiment 1, we manipulated both color and the (go/stop) meaning of the stimuli. Results showed the expected interaction between type of word and color both for hit rates and response times. Overall, we observed that participants were faster and more accurate to evaluate go-related words when the stimuli were presented in green. In the case of stop-related words, this pattern was reversed (i.e., slower responses and more errors) when the stimuli were presented in green.

Previous research has mainly focused on the color red (for an exception, see Lichtenfeld, Elliot, Maier, & Pekrun, 2012), including green as the control color (e.g., Gnambs et al., 2010). A potential issue with this approach is that it not possible to ascertain whether red is inhibiting performance or if is green that facilitates it. In our experiments, besides testing the colors green and red, we have included gray as the control. The overall pattern observed with gray falling in between the green and red conditions actually suggests facilitation and inhibition effects.

Moreover, and extending previous studies (e.g., Mentzel et al., 2017), we also included a set of words unrelated to movement (i.e., neutral words). Regarding the categorization of these words, we observed a higher proportion of go responses as well as faster reaction times when the neutral words were presented in green in comparison to those presented in red. This pattern of findings suggests than, even when semantic content is not informative for the categorization (i.e., absence of a clear stop or go meaning), then color was used as a cue to complete the task.

Experiment 2 constitutes a critical test to the functional autonomy acquired by these colors in shaping task performance. Specifically, in this experiment, both the stimuli (words and pseudo-words) and the task (lexical decision) were not related to typical meanings associated with red or

green (e.g., stop and go). In line with the results of the categorization of neutral words observed in Experiment 1, the results showed that color directly influenced accuracy and speed in a lexical decision task, with higher hit rates as well as faster reaction times observed in response to neutral words presented in green than in red. For pseudo-words, color did not seem to affect accuracy but only response time. In contrast with words, the classification of pseudo-words presented in green was slower.

The effect of color observed for neutral words seems to suggest that green facilitates performance independently of semantic content. However, the pattern of response times observed for pseudo-words challenges this conclusion. This interaction between type of stimuli and color, suggests a boundary condition that is not related to the color associations with the stimuli but with the nature of the task. One possible explanation is that when classifying a pseudo-word people are providing a "negation" (i.e., NOT a word). As green and red may also ground meaning associated with affirmation / negation, providing such a response to a stimuli presented in green might constitute an interference situation. This remains a question for further examination.

A potential shortcoming of the current research was that we only used words as stimuli. Future studies could use the paradigm used by Pravossoudovitch et al. (2014), in which the colors are used as context for the categorization of symbols (e.g., Prada, Rodrigues, Silva, & Garrido, 2016). Moreover, the only associations examined here were the green/go, and red/stop. Future studies using a similar paradigm could further examine the effects of color using other known associations (e.g., safety/danger). Finally, other paradigms have the potential to clarify such color effects, namely priming paradigms with colors as primes and words as targets.

Overall, our findings support the idea that contextual cues such as colors can shape cognitive processes and behavior. Moreover, in experiment 1 we extended the association of red and green to other concepts (e.g., other than danger and safety or failure and success). This is particularly relevant because color (like other cues) can ground more than one concept, much like verticality grounds valence (e.g., Meier & Robinson, 2006) and power (e.g., Schubert, 2005) and horizontality

grounds time (e.g., Lakens, Semin, & Garrido, 2011), and politics (Farias, Garrido, & Semin, 2013). In experiment 2, which constitutes a more innovative approach, we examined the effect of color in the absence of relevant semantic associations, illustrating a novel instance of functional autonomy. However, and despite some evidence for the functional autonomy of color - green speeds up and red slows down performance in the classification of neutral words, this pattern was not observed for pseudo-words. This interaction suggests that these colors should not be used as behavioral cues irrespective of the contexts or specific associations they maintain.

Finally, the reported findings may be informative for applied domains. For example, in the eating behavior domain, it has been shown that red can act as a subtle stop cue helping the consumer to prevent incidental snack food and soft drink intake (Genschow, Reutner, & Wänke, 2012) and that green labels are perceived as signaling healthfulness cues even when presented on candy bars (Schuldt, 2013). Based on the current findings, it would be interesting to examine for example if green could actually be used to promote the consumption of healthful foods or drinks and red to limit the purchase and consumption of unhealthy ones.