Can the "Strength of Connections" Account Explain Picture and Word Naming and Categorization Data?

by

Sherif Soliman

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada 2015

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Potter and Faulconer (1975) reported that participants were faster to read words aloud than they were to name pictures, but were faster to categorize pictures than words. They took this as falsification of an account in which picture representations have to be translated to their verbal counterparts before semantic information about the object could be activated, and corroboration of an account that claimed that both picture and verbal representations have direct access to semantics. Their direct access account does sufficiently explain the findings, but there are other plausible accounts that could better explain the specific pattern they found. In the present work, I sought to explain their findings using an alternative account proposed by Besner et al. (2011), which posits that the strength of connections between localist modules vary as a function of use. Besner et al. (2011) found that participants were faster to perform parity judgments on Arabic numerals than on number words, but were equally fast at reading/naming aloud both, a finding that cannot be explained by the received view that frequency effects exist as resting levels of activation in localist modules, but *can* be explained by the strength of connections (SOC) account. The present experiments replicate Potter and Faulconer's (1975) findings, and also replicate findings by Rogers and Monsell (1995) that show that switching from one task to the other incurs an RT cost. I derived and predicted that these switch costs would be larger in blocks in which the task instructions conflict with the task participants have learned to associate with a stimulus (default set), than in blocks in which task instructions are compatible with the default set. I find data in support of this prediction for words, but not for pictures. Possible reasons for the failure to find the predicted pattern for pictures are briefly discussed.

Acknowledgments

I am grateful to my supervisor, Dr. Derek Besner, for his invaluable input, knowledge, and direction, without which this thesis would not have been completed.

I am also grateful to my thesis readers, Drs. Jonathan Fugelsang and Evan Risko, for their thoughtful and constructive comments.

Finally, I would like to thank Tanya Jonker for her valuable comments, discussion, and encouragement, and acknowledge the Natural Sciences and Engineering Research Council (NSERC) and the Ontario Student Assistance Program (OSAP) for their financial support of my Masters work in the form of postgraduate scholarships.

Table of Contents

List of Tables

List of Figures

Accounts of performance in various tasks often start by specifying a variety of mental representations (modules), each specialized for a particular information processing job (see Figure 1 for an example). These representations are then linked in various ways with each other, and, importantly, have a particular order and certain specifications regarding the directions of activation and communication between them. The combination of these representations and connections provide a flow-chart way of thinking about how a task is performed.

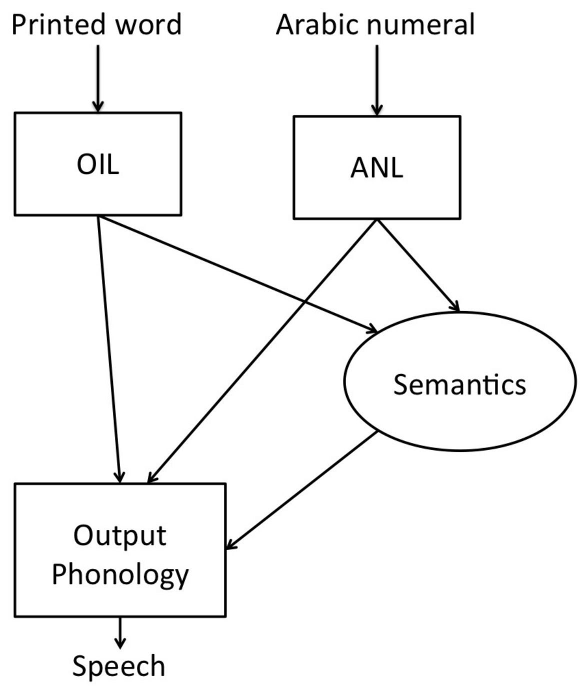

Figure 1. A localist framework with distinct modules for written words (Orthographic Input Lexicon, OIL) and Arabic numerals (Arabic Numeral Lexicon, ANL) (Besner et al., 2011).

To illustrate, consider a localist framework of how a single number word (e.g., two) or Arabic numeral (e.g., 2) might be read aloud. In the example from Besner et al. (2011) provided in Figure 1, the two modules, OIL (Orthographic Input Lexicon) and ANL (Arabic Numeral Lexicon) are two distinct lexicons that process only printed words and Arabic numerals respectively. The Output Phonology Module contains entries for all the phonological codes associated with the inputs. The Semantic module contains conceptual information related to those inputs, such as numerical magnitude (ordinality), and parity (whether the number is odd or even).

The arrows between the modules reflect the flow of activation. For example, activation spreads from the OIL and ANL directly to Output Phonology, and first through Semantics.[1]

It is important to note that, in the standard and well-received view (see the Dual-Route Cascaded model of visual word recognition and reading aloud, (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001)), the arrows between the different modules denote a binary status of whether a connection exists or not, and the directionality of this connection, nothing further. Any cognitive phenomena that these models attempt to explain, such as the frequency effect in which words that are seen more frequently are processed faster than words seen less frequently, is only explained in terms of what happens inside a module, regardless of whether they are connected to other modules, or the directionality of that connection. With the exception of directionality, all connections are equal in strength. With this view of localist models in mind, I will now introduce prior findings that inform the present work.

Picture and word naming and categorization

Potter and Faulconer (1975) report an experiment in which they have people read or name aloud words and pictures, or perform a semantic classification on them. Participants were presented with 96 items, half as words and half as pictures, in alternating blocks of 16 items each. In one block they were to read or name aloud the words or picture that were presented, and in the other block the researcher would specify a category before presentation, and the participant's task was to respond with "Yes" or "No" to whether an item was a member of that category.

Potter and Faulconer's motivation was to pit two accounts of memory representation against each other. The first account, which I will refer to as the "direct activation" account, posited that when an object (e.g., chair) is presented in different forms, say as a picture, or a written word, these two representations directly activate a common abstract representation of the concept of "chair" (semantics). The second account, which I refer to as the "verbal mediation" account, posited that only the written representation of the word was connected to the abstract

[1] In the Besner et al. paper, the connection between Semantics and Output Phonology is bidirectional, however since directionality is irrelevant to the current work, I present a simplified version of the model in which activation travels only from Semantics to Output Phonology.

representation of a chair, and that the picture of the object must first be translated into a verbal representation before abstract information about the object can be activated. This process of verbal mediation takes time, which is why the second claim could explain why it took longer to name a picture of an object than it took to read aloud its orthographic representation.

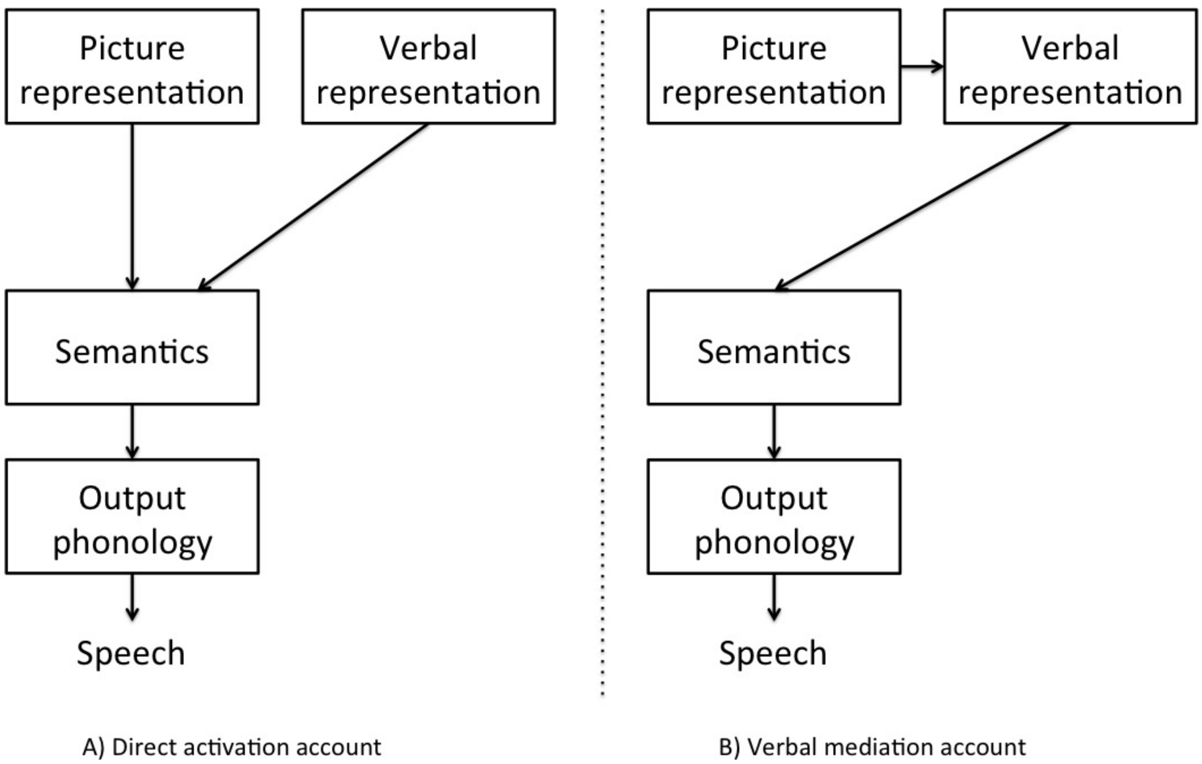

Figure 2. The two models tested by Potter and Faulconer (1975). A) The direct activation model, which posits that written word and picture representations of an object each directly activate one abstract representation for the object in semantics. B) The verbal mediation account which posits that a picture representation must be translated to its verbal counterpart before the semantic information for the object can be activated.

Potter and Faulconer (1975) found that participants were slower to name a picture than they were at categorizing it. In contrast, they found that words were faster to read aloud than the pictures were to name.

The verbal mediation account predicts that participants would be slower to respond to pictures than words regardless of the operation, since the picture had to be translated to its verbal representation before semantic information could be activated. As a result, Potter and Faulconer

(1975) took their findings as falsification of the verbal mediation account and consistent with the direct activation one.

Potter and Faulconer's findings are sufficient to falsify the verbal mediation account, and in the contest between that account and the direct access account, to support the latter. However, it was not necessary for them to obtain the crossover pattern of results they did to favour the direct access account. For example, demonstrating any single task in which responding aloud to pictures was not slower than responding to words would have been sufficient for their theoretical conclusions. Therefore, while their findings were sufficient to adjudicate between the two accounts they were considering, I am interested in explaining the specific crossover interaction they observed. Specifically, why would words be faster to read than they are to categorize, with the reverse being true for pictures?

A strength of connections account

There is, however, a different account that may be useful in considering Potter and Faulconer's findings. A localist account similar in structure to the one outlined earlier, of how participants performed in Potter and Faulconer's experiments assumes that words are processed via the Orthographic Input Lexicon (OIL), while pictures are processed through a separate lexicon for pictorial representations (which I will call a Pictorial Lexicon, PL; see Seymour, 1979, for a discussion of a pictogen system analogous to the Pictorial Lexicon I assume in our model). Once recognized, either the phonological code for the object, or semantic information related to its category would need to be activated, for reading aloud/naming or categorization tasks respectively.

Such an account was proposed by Besner et al. (2011), who posited that, instead of the connection between two modules having a binary state of either existing or not, the strength of these connections can vary over time as a function of the frequency with which they are used. They tested this account by measuring RTs for naming or reading-aloud Arabic numerals and their alphabetic counterparts. Specifically, it is the received view that frequency effects (i.e. that high frequency words are faster to read aloud than low frequency words) arise due to differences in resting levels of activation in localist modules; modules such as the ones containing entries for Arabic numerals and words (including number words). This view predicts that, since Arabic numerals appear more frequently than their corresponding number words, Arabic numerals should be faster to respond to compared to number words regardless of what the task is. On the other hand, an account that posits that these frequency effects result, at least partially, from the varying strengths of connections between stimulus-specific (e.g., OIL) or task-specific (e.g., Semantics) modules would predict an interaction of stimulus and task in RT. In other words, RT would not only be affected by resting levels of activation in the Arabic or orthographic lexicons, but also by the strength of connections between those lexicons and semantics, or the POL. Besner et al. (2011) found that performing parity judgments (determining whether the number is odd or even) was faster for Arabic numerals than number words, but there was no difference in RT for naming/reading aloud either stimulus type. This result contradicts any accounts that explain frequency effects as resulting from varying resting levels of activation in modules alone, and supports an explanation involving the strength of connections between these modules.

This strength of connections (SOC) account provides an alternative explanation for Potter and Faulconer's results. It is reasonable to assume that producing the phonological code for a written word is a more frequent "task" than producing specific semantic information about it. Also reasonable is the assumption that producing semantic information about a picture, or producing a phonological response reliant on semantic information in a picture, is a more common action than naming the represented object aloud. If one takes these stimulus-task frequencies into account, I would expect that if the connections between the different modules are strengthened as a result of their frequent use, then the connection between the OIL to the phonological output lexicon (POL) would be stronger than the connection between the OIL and semantics. I would also expect that the connection between the PL and semantics to be stronger than the one between the PL and the POL.

In an experimental context, I would expect these varying connections strengths to manifest themselves in the time taken to perform a particular task. The stronger the connection between modules, the faster I would expect a response to be produced. Specifically, this account can accommodate Potter and Faulconer's results. If a stronger connection between modules would translate into faster responses, and, as conjectured earlier, one would expect that for words the connection between OIL and POL to be stronger than the one between OIL and Semantics, with the reverse being true for pictures, then one would expect that words would be faster to read aloud than they are to categorize, whereas pictures would be faster to categorize than they are to name aloud.

The present investigation

The purpose of the present work was to a) show how this SOC account can accommodate Potter and Faulconer's data, and b) derive and test a new prediction. I therefore now turn to work by Rogers and Monsell (1995) who examined the costs of predictable task-switching in the context of some simple cognitive tasks, and provide an experimental paradigm that I will use in the present studies.

Rogers and Monsell (1995) examined the costs of switching between tasks that are afforded by the same stimulus, and used their data to discuss some notions of cognitive control. Cognitive experimental paradigms very frequently employ very specific, and to the participants arbitrary-seeming, task structures and instructions. To make inferences about underlying cognitive mechanisms, cognitive psychologists typically assume that participants are accurately following instructions in a consistent fashion. More specifically, researchers can easily rely on the fact that people can adopt task instructions at will, maintain those instructions in memory, and apply them successfully when necessary. Rogers and Monsell refer to this "[adoption] of task-sets at will" as "endogenous control".

Rogers and Monsell visualize the adoption of this task-set as binding components (modules or processes) that will form a chain to accomplish the desired task. For example, when the participant is asked to read aloud words, they "know" to emphasize the OIL to the POL connection, whereas if they knew they are instructed to categorize those words, they would emphasize the connection between OIL and semantics, which is then connected to whichever module is necessary for the required response.

However, this ability to maintain and follow specified and arbitrary-seeming task instructions is not seamless. In addition to the instructions presented by the task, Rogers and Monsell discuss the tendency of a stimulus to evoke a certain action; they call that *exogenous* control. Rogers and Monsell do not further explore this *exogenous* control idea in their experiments, but it is a concept that is at the heart of the work reported here.

> ... stimuli can of themselves activate or evoke in a person a tendency to perform actions (or tasks) habitually associated with them, irrespective of prior intention, and sometimes in conflict with prior intention. We refer to this as *exogenous* control.
>
> (Rogers & Monsell, 1995, p. 208)

Rogers and Monsell (1995) use a predicable AABB pattern of tasks. In other words, if the experiment involves two tasks, A and B, the trials are presented such that the participant will perform task A on two successive trials, and then perform task B on the subsequent two trials. Therefore, at any given moment, the participants knows what the next task was going to be. Rogers and Monsell were interested in the effects that this predictability would have on performance. Throughout their experiments, Rogers and Monsell found that switching between two tasks afforded by the same stimulus produced reliable costs in performance in terms of RT and errors. For example, in a task that involved participants either deciding whether Arabic numerals as being odd or even, or whether a letter was a consonant or a vowel, Rogers and Monsell (1995) found that both mean RTs and error rates were reliably greater in switch trials (trials that followed a trial with a different task) than nonswitch trials.

They also found that increasing the Response-Stimulus (R-S) interval to over a second resulted in a reduction of this cost in RT, but a significant portion of this cost remained. Critically, they also found that:

> When the stimulus contained attributes associated with the competing task, both performance on nonswitch trials and the efficiency of switching were impaired. We interpret this as a manifestation of the tendency of stimuli to activate task-sets associated with them – a task-cuing effect.
>
> (Rogers & Monsell, 1995, p. 228).

It is important to elaborate on the final point. What do Rogers and Monsell mean by the stimulus having an "attribute associated with the competing task"? It is already the case that the stimuli afford more than one task, and in fact can be processed in at least two different ways, both of which are in the set of tasks in the experiment. What they mean is that, prior to the presence of a task-set specified by the experiment, there existed a "default-set" that, in the absence of specific instructions, encourage one to perform a specific task in the presence of a specific attribute in the stimulus.

In other words, participants have learned over time that certain stimuli (e.g., printed words) are most strongly associated with certain actions (e.g., reading aloud). This is what Rogers and Monsell refer to as "endogenous control", and it is what produces interference that takes the form of increased cost in task-switching when the task is not what the stimulus is strongly associated with, relative to when the task is the one prompted by the stimulus.

In our terms, the default set exists as the learned strengths of connections between different modules. As explained earlier, this explanation can account for Potter and Faulconer's findings regarding reading/naming aloud words and pictures. That is, as can be seen in Figure 3, it is assumes that the connection between the OIL and the POL to be stronger than that between OIL and semantics, which translates to faster reading aloud of words than categorization of them. Conversely, I assume that the connection between the PL and semantics to be stronger than that between PL and the POL, which translates to faster categorization of pictures compared to naming them aloud.

In the present work I use Rogers and Monsell's AABB paradigm to test this SOC account. Since the stimuli afford more than one task, when the task-set (the one specified by the experiment) coincides with the participant's default set, I consider this to be a **compatible** task-set. For example, reading words aloud, or activating specific semantic information about pictures are examples of compatible trials. In contrast, when the task-set conflicts with the default set, I consider that an **incompatible** task-set. For example, activating specific semantic information about a word, or naming the object in a picture aloud are examples of incompatible

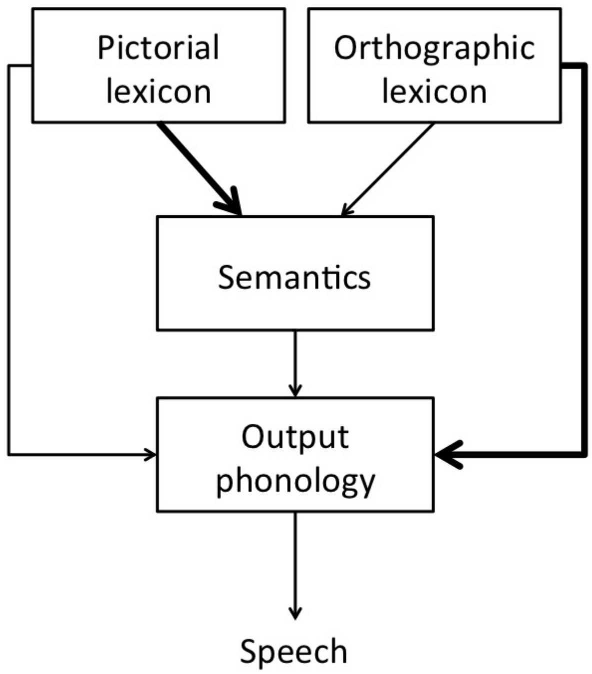

Figure 3. The SOC account proposed by Besner et al. (2011), as applied to picture and word processing, posits that the connection between the Orthographic Input Lexicon (OIL) and Phonological Output Lexicon (POL) is stronger than that between OIL and Semantics, and that conversely, that the connection between Pictorial Input Lexicon (PIL) and Semantics is stronger than that between PIL and OIL.

trials.

Given that, I make the following predictions:

1. Words will be faster to read aloud than they are to categorize, and conversely, images will be faster to categorize than they are to name aloud.

   This prediction follows from what I expect the default set to be in terms of strength of connections between modules, and what that would translate to in terms of performance. It is reasonable to expect that the phonological output codes of words are more often directly activated than specific semantic information about those words, and reasonable to expect that semantic information of images are activated more frequently than the phonological output codes for the objects represented in the image. As a result, the connection between the OIL and POL should be stronger than that between OIL and semantics, leading to faster reading aloud of words than categorization. The connection between the PL and semantics would be stronger than that between PL and POL, leading to faster categorization of images than naming aloud.

2. Responses on trials that consist of a task that is different from the one on the previous trial (**switch** trials) should be slower than responses on trials that maintain the same task as the prior trial (**stay** trials). That is, as in Roger and Monsell's (1995) findings, I expect that participants should incur a cost when they have to reconfigure their planned actions given a change in the required task even when they can predict the pattern of tasks ahead of time.

3. Critically, the switch costs should be larger for **incompatible** blocks than **compatible** blocks. That is, in addition to the baseline cost of rearranging planned actions for trials that signal a change of task to be performed, I expect that there should be an added cost associated with having to override the default set when the instructed task is incompatible with what the default set indicates. For example, participants should be more slowed if they are required to categorize a word on a trial that follows one in which they had to name an object in a picture aloud, than if they had to read aloud a word on a trial that follows one in which they had to categorize a picture.

Experiment 1

Method

Participants. Twenty-nine undergraduate students from the University of Waterloo received one credit towards a psychology course of their choice in return for participation. Due to technical errors in the experiment program, data from 4 participants was excluded, leaving 25 participants whose data is included in the analyses below.

Materials. The stimuli used in this experiment were obtained from Snodgrass and Vanderwart (1980), and were used in the subsequent experiments. A total of 200 pictures were selected from their set, such that 100 were classified as *animate* and the other 100 classified as *inanimate*.[2]

The names associated with the pictures were used as the list of words corresponding to the pictures.

Out of the total set of 100 pictures and word pairs (totaling 200 words and pictures), 20 were used for practice trials, leaving 80 pairs that were used in all experimental blocks for all reported experiments.

Stimuli for all experiments were presented using E-Prime 2.0 software running on a 22" Flatron LCD display at a 1680 $\times$ 1050 resolution. Words were displayed in 36 pt Courier New lowercase font, and pictures were displayed at their native dimensions of 300 $\times$ 300 pixels.

Procedure. To ensure that participants were familiar with the pictures and their corresponding names, they were presented with each picture and its corresponding name below it one by one, and instructed to familiarize themselves with those pairs. They were told they did not need to memorize them, only be familiar with them, and were allowed to set the pace of the presentation by pressing the space bar to continue to the next picture-word pair until they had surveyed all 200 pairs.

The experiment consisted of two blocks. In the compatible block, participants were instructed to respond to stimuli in accordance to what the default set is considered to be. In other words, if

[2]Participants were instructed than an object was animate if it belonged to the animal kingdom (e.g., humans, animals, and insects).

participants saw a word they were to read it aloud, and if they saw a picture they were to respond with "Yes" to indicate an object being animate, or "No" to indicate an object being inanimate. In the incompatible block, the task-stimulus association was reversed: if they saw a picture, participants were to name it aloud, and if they saw a printed word they were to respond with "Yes" to indicate the object described by the word being animate, or "No" to indicate an object being inanimate.

The order of compatible and incompatible blocks was counterbalanced across subjects, and in both blocks participants were told that the words and pictures would be presented in a "predictable AABB pattern", and were given instructions on how to respond to each.

Each experimental block began with 20 practice trials, after which, participants were prompted to ask the researcher any questions they wanted, and then proceeded to the 40 experimental AABB sets, totaling 160 experimental trials.

On each trial, a fixation sign consisting of a + appeared in the centre of the screen for 400 ms, after which the word or picture appeared on the screen until the participant responded, or 4000 ms elapsed, whichever came first. After the voice key registered the participant's vocal response, a research assistant coded whether the participant responded accurately or not, and whether a microphone or vocal error had occurred, indicating a mistrial. As a result of the need for manual accuracy coding by a research assistant, the intertrial interval (ITI) varied from trial to trial.

Results and discussion

Error trials comprised 1.7%, and mistrials[3] 5.8% of all experimental trials. Therefore, a total of 7.5% of trials was excluded from further analysis.

A trial by trial outlier analysis was the performed on correct responses, and all trials with RTs over 2.5 SDs away from the mean, within subject and condition, were removed (Van Selst & Joliceur, 1994) . Outlier analysis for subject grand mean RTs and accuracy rates over 2.5 SDs away from the mean did not exclude any participants. Overall, 3% of trials were considered RT outliers and were removed from further analysis.Table 1 shows the mean RTs and % errors for

[3]Mistrials represent voice-key errors, unintentional noise by the participant, or other background noise.

each condition.

Table 1
Mean RTs (ms) with % errors in parentheses as a function of compatibility (compatible vs. incompatible), stimulus type (words and pictures), and switch vs. stay conditions. For each stimulus type, notation above mean RT explains the kind of trial. For example, P $\rightarrow$ W *indicates that the current word trial was preceded by a picture trial.*

	Compatible				Incompatible			
Stimulus	Switch	Stay	±CI	**Cost**	Switch	Stay	±CI	**Cost**
Words	P $\rightarrow$ W	W $\rightarrow$ W			P $\rightarrow$ W	W $\rightarrow$ W		
	580 (.4%)	511 (.1%)	34.13	**69**	936 (4%)	775 (3.2%)	34.13	**161**
Pictures	W $\rightarrow$ P	P $\rightarrow$ P			W $\rightarrow$ P	P $\rightarrow$ P		
	756 (.1%)	644 (.1%)	34.13	**112**	865 (3%)	826 (3%)	34.13	**39**

Omnibus analysis of RT. The data was analyzed as a 2 $\times$ 2 $\times$ 2 repeated-measures analysis of variance (ANOVA), with compatibility (compatible vs. incompatible), stimulus type (words and pictures), and switch (switch vs. stay) conditions as factors.

There was a main effect of compatibility, in which overall RTs for incompatible trials were slower than those for compatible trials, $F(1, 24) = 147.75$, $MSE = 17424.84$, $p < .001$, $\eta^2 = .86$, a main effect of stimulus type, in which RTs for pictures were slower than RTs for words, $F(1, 24) = 80.18$, $MSE = 3331.94$, $p < .001$, $\eta^2 = .77$, and a main effect of switching such that RTs were slowed when the current trial contained a different stimulus type than when the current trial repeated the same stimulus type, $F(1, 24) = 61.87$, $MSE = 7457.97$, $p < .001$, $\eta^2 = .72$. There was a significant compatibility by stimulus interaction, $F(1, 24) = 24.87$, $MSE = 13886.76$, $p < .001$, $\eta^2 = .51$, and a stimulus by switch/stay interaction, $F(1, 24) = 26.66$, $MSE = 782.73$, $p < .001$, $\eta^2 = .53$. Finally, the overall three-way interaction was significant, $F(1, 24) = 24.06$, $MSE = 3716.34$, $p < .001$, $\eta^2 = .50$.

Errors. Participants made more errors in the incompatible block compared to the compatible block, $F(1, 24) = 22.03$, $p < .001$. Neither the threeway interaction, nor any of the two-way interactions were significant in a repeated-measures ANOVA of the errors.

Predictions. I now discuss the specific predictions outlined in the introduction.

Prediction 1: Words should be faster to read aloud than they are to categorize, and inversely, pictures should be faster to categorize than they are to name aloud.

To test this prediction, I conducted t-tests on the means for reading/naming aloud and categorization for word and picture data separately. The t-tests confirm the prediction: words were faster to read aloud (M = 546) than they were to categorize (M = 853), $t(24) = 12.49$, $p < .001$, and pictures were faster to categorize (M = 702) than they were to name aloud (M = 846), $t(24) = 5.78$, $p < .001$.

Prediction 2: Responses on switch trials should be slower than responses on stay trials.

This prediction was already tested and confirmed in the omnibus ANOVA with the significant main effect of switch showing that RTs were slowed when the current trial contained a different stimulus type than when the current trial repeated the same stimulus type, $F(1, 24) = 61.87$, $MSE = 3716.34$, $p < .001$.

Prediction 3: Switch costs should be larger for incompatible blocks than in compatible blocks.

The mean switch costs can be seen in Table 1. The compatibility × switch interaction in the omnibus ANOVA was not significant, implying no difference of switch costs as a function of compatibility. The threeway interaction of compatibility × stimulus type × switch was significant. However, conducting a t-test comparing switch costs for compatible vs. incompatible blocks does not yield a significant difference, $t(24) = .88$, $p = .39$.

The explanation for this can be found by examining the two-way interaction separately for words and pictures. For words, I do find that switch costs are larger in the incompatible block than in the compatible block, $t(24) = 5.05$, $p < .001$. However, for pictures, switch costs were larger in the compatible block than they were in the incompatible block, $t(24) = 3.51$, $p < .01$. I elaborate on these findings further in the discussion section.

Experiment 2

The results of Experiment 1 replicated Potter and Faulconer's (1975) findings in that words are faster to read than they are to categorize, but pictures were faster to categorize than to name. I also replicated Rogers and Monsell's (1995) finding that there are reliable costs in terms of RT on switch trials that change the stimulus and task from the previous trial.

The results of Experiment 1 include a counter-intuitive finding. I predicted that switch costs should be larger in incompatible blocks than in compatible blocks. That prediction was confirmed for words, but reversed for pictures. It is unclear why the pattern was different for words than for pictures. One possibility is that, given the smaller set of possible "tasks" to perform on a word compared to the ambiguity regarding what to do with an pictures, that the predictable AABB of trials provides different advantages in preparation for words and pictures.

To avoid these possible complexities of a predictable pattern of trials, I discarded the AABB structure for Experiment 2. Now, words and pictures were presented in a random (mixed) order within each block. All other methods and procedures are the same as Experiment 1.

Method

Participants. Twenty-eight undergraduate students from the University of Waterloo received one credit towards a psychology course of their choice in return for participation.

Materials and procedure. The same stimulus set and materials from Experiment 1 were used.

Participants were shown the same words and pictures from Experiment 1 in a different random order for each subject, and were told to respond according to the instructions for compatible and incompatible blocks outlined in Experiment 1. That is, they were to read words aloud and categorize pictures in the compatible block, and categorize words and name objects in pictures aloud in the incompatible block.

All other materials and procedures were the same.

Results and discussion

Error trials comprised 2.1%, and mistrials 6.3% of all experimental trials. A total of 8.4% of all trials were therefore excluded from further analysis.

A trial by trial outlier analysis was then performed on correct responses, and all trials with RTs over 2.5 SDs away from the mean, within subject and condition, were removed (Van Selst & Joliceur, 1994) . Like in Experiment 1, outlier analysis for subject grand mean RTs and accuracy did not exclude any participants. Overall, 2.9% of trials were considered RT outliers and removed from further analysis. Table 2 shows the mean RTs (in ms) for each condition.

Since trials were presented in a random order instead of a predictable AABB pattern, trials were coded as switch or stay in the data analysis stage. A trial was coded as a switch trial if it instructed the participant to perform a different task as the previous trial, and was coded as a stay trial if it instructed the participant to perform the same trial as the previous trial.

Table 2
Mean RTs (ms) with % errors in parentheses as a function of compatibility (compatible vs. incompatible), stimulus type (words and pictures), and switch vs. stay conditions. For each stimulus type, notation above mean RT explains the kind of trial. For example, P $\rightarrow$ W *indicates that the current word trial was preceded by a picture trial.*

	Compatible				Incompatible			
Stimulus	Switch	Stay	±CI	**Cost**	Switch	Stay	±CI	**Cost**
Words	P $\rightarrow$ W	W $\rightarrow$ W			P $\rightarrow$ W	W $\rightarrow$ W		
	672 (.8%)	581 (.2%)	70.05	**91**	1065 (4.6%)	896 (2.7%)	70.05	**169**
Pictures	W $\rightarrow$ P	P $\rightarrow$ P			W $\rightarrow$ P	P $\rightarrow$ P		
	853 (.6%)	790 (.6%)	70.05	**63**	1034 (3.7%)	975 (3.5%)	70.05	**59**

Omnibus analysis of RT. The data was again analyzed as a $2 \times 2 \times 2$ repeated-measures ANOVA, with compatibility (compatible vs. incompatible), stimulus type (words vs. pictures), and switch (switch vs. stay) conditions as factors.

There was a main effect of compatibility, in which overall RTs for incompatible trials were slower than those for compatible trials, $F(1, 27) = 92.78$, $MSE = 43889.04$, $p < .001$, $\eta^2 = .77$, a main effect of stimulus in which overall RTs for pictures were slower than RTs for words, F(1,

27) = 86.30, *MSE* = 7101.79, $p < .001$, η^2 =.75, and a main effect of switching the stimulus or task from the previous trial such that RTs were slowed when the current trial contained a different stimulus type than when the current trial repeated the same stimulus type, $F(1, 27) = 94.21$, *MSE* = 4593.39, $p < .001$, $\eta^2 = .79$. There was a significant compatibility by stimulus type interaction, $F(1, 27) = 45.20$, *MSE* = 10176.32, $p < .001$, $\eta^2 = .61$, a marginally significant compatibility by switching interaction, $F(1, 27) = 4.35$, *MSE* = 4426.71, $p = .047$, $\eta^2 = .14$, and a significant stimulus type by switching interaction, $F(1, 27) = 36.42$, *MSE* = 1666.55, $p < .001$, $\eta^2 = .58$. Finally, the overall three-way interaction was significant, $F(1, 27) = 4.94$, *MSE* = 4654.32, $p = .035$, $\eta^2 = .15$.

Errors. Participants made more errors in the incompatible block compared to the compatible block, $F(1, 27) = 26.14$, $p < .001$, and made more errors on switch trials than on stay trials, $F(1, 27) = 8.78$, $p < .01$. Additionally, the stimulus type $\times$ switch vs. stay interaction was significant, $F(1, 27) = 6.6$, $p < .05$. Specifically, for pictures there were no % error differences between switch and stay trials, $t(27) = .37$, $p > .05$, however, for words participants made more errors on switch trials than on stay trials, $t(27) = 3.16$, $p < .01$.

Predictions. As with Experiment 1, I now outline tests for specific predictions outlined in the introduction.

Prediction 1: Words should be faster to read aloud than they are to categorize, and inversely, pictures should be faster to categorize than they are to name aloud.

To test this prediction, I conducted t-tests on the means for reading/naming aloud and categorization for word and picture data separately. The t-tests confirm the prediction: words were faster to read aloud (M = 628) than they were to categorize (M = 981), $t(27) = 11.62$, $p < .0001$, and pictures were faster to categorize (M = 822) than they were to name aloud (M = 1004), $t(27) = 5.64$, $p < .0001$.

Prediction 2: Responses on switch trials should be slower than responses on stay trials.

This prediction was already tested and confirmed in the omnibus ANOVA with the significant main effect of switch showing that RTs were slowed when the current trial contained a different stimulus type than when the current trial repeated the same stimulus type, $F(1, 27) = 94.21$, $MSE = 4149.59$, $p < .001$.

Prediction 3: Switch costs should be larger for incompatible blocks than in compatible blocks.

The mean switch costs can be seen in Table 2. Unlike in Experiment 1, the compatibility × switch interaction in the omnibus ANOVA was marginally significant. The threeway interaction of compatibility × stimulus type × switch was significant. Unlike in Experiment 1, conducting a t-test comparing switch costs for compatible vs. incompatible blocks does show that switch costs were larger for incompatible than compatible trials , $t(27) = 2.11$, $p = .04$.

Once again, the pattern of results is different for words and pictures. When analyzing word data separately, I find that switch costs are larger in the incompatible block than in the compatible block, $t(27) = 2.93$, $p < .01$. However, and unlike in Experiment 1, there is no difference in switch costs between the compatible and incompatible block for pictures, $t(27) = .11$, $p = .91$.

General discussion

The present work set out to a) replicate Potter and Faulconer's (1975) findings on the time it takes to read/name and categorize words and pictures, and b) Rogers and Monsell's (1995) work on the costs of switching between simple cognitive tasks. It also set out to c) explore the SOC account as applied to picture and word naming and categorization.

Potter and Faulconer (1975) found that words were read fast than pictures (see Cattell, 1886 for the earliest demonstration of this effect), whereas pictures were faster to categorize than words. Besner et al. (2011) proposed an account in which frequency effects are explained not just in terms of resting levels of activation inside localist modules, but also in the strength of the connections between various modules. The account proposes that the strength of connections between modules (e.g., from the OIL to POL) are formed as a function of the frequency with which activation spreads from one module to the other, and that these varying strengths of connections manifest in terms of RT, where the stronger the connection, the faster the RT. Specifically, since I expect that words are more often read aloud than categorized, and that people activate semantic information about a picture more often than they name the object pictured, the SOC account predicts Potter and Faulconer's results (1975). Indeed, the present work replicated the finding that words were faster to read aloud than they were to categorize, while the reverse was true for pictures.

The SOC account makes a further prediction related to the costs of switching between tasks within an experiment. Rogers and Monsell found that switching from one task to another when both tasks were afforded by the same stimulus (e.g., switching from reading a word aloud to categorizing a word) incurs a cost both in terms of errors and RT. They propose that this cost is due to the need to episodically bind the modules required to perform the instructed task on the provided stimulus (e.g., to switch from reading words aloud to categorizing them, a participant would have to "disconnect" or disengage the OIL from the POL, and instead emphasize activation through Semantics to retrieve the categorical information required), and that this rearrangement takes time. Importantly for the present work, in addition to this adoption of the task-set, which

Rogers and Monsell term *endogenous* control, Rogers and Monsell discuss the idea that independently of the task-set, a stimulus is likely to evoke a certain action, which they term *exogenous* control (Rogers & Monsell, 1995).

I refer to this idea of exogenous control as the "default set": the action a certain stimulus would evoke by default in the absence of a task-set. In a localist model context, this default set can be thought of as the path with the strongest connections for each stimulus type (e.g., reading aloud for words: OIL $\rightarrow$ POL; semantic information retrieval for pictures: PL $\rightarrow$ Semantics $\rightarrow$ POL).

With this in mind, I derived our second and third predictions. I predicted that I would find a reliable switch cost when participants had to switch from one task to the other. Indeed, in both experiments RTs are significantly slowed both for words and pictures when the current trial dictated a task different from the previous trial [4].

Critically, I also predicted that the observed switch costs would interact with the effect of the task set conflicting with the default set. Specifically, I predicted that the cost for switching tasks would be larger when the task set is incompatible with the default set, compared to when the task set agrees with the default set. In both experiments reported here, this prediction is confirmed for words, but not for pictures.

There is no immediately obvious post hoc explanation for why the results for pictures are inconsistent with our account. That said, one possible explanation is that the range of semantic information evoked by a picture is much larger than the ones afforded by a word, and therefore that the strength and identity of the default set are more difficult to determine than they are for words. However, the fact that pictures are categorized faster than they were to name aloud corroborates the expectation that activation of semantic information is the default set, or at least more of a default set than naming the pictures aloud. Additionally, it is also unclear why the

[4] One possible point that could be raised is the possibility that the slowed RTs that I interpret as a cost for task-switching could actually be a cost for stimulus switching – in other words, task and stimulus are confounded. However, in a third experiment that I do not discuss here, stimulus type was blocked while the task changed in a predictable AABB pattern. This mirrors a more traditional task-switching paradigm in which the task changes while the stimulus is held constant. A reliable switch cost of comparable magnitude to the one found in present experiments was observed for both words and pictures, indicating that the original cost observed was associated with switching the task, not stimulus type.

pattern for pictures changes when the experiment shifts from a predictable AABB pattern of tasks to a randomly mixed one, while the pattern for words remains the same.

These data may mean that the strength of associations account is incorrect or incomplete, and cannot make predictions for all stimuli types. Perhaps a strength of associations account can only explain findings for words and numbers (see Besner et al., 2011) that rely on processing of orthography, whereas the processing of pictures is different in a critical, but as of yet not understood way that interacts differently with task switching.

Conclusion

The current work tested the ability of the SOC account to explain Potter and Faulconer's (1975) findings on the speed to read/name aloud or categorize words and pictures, and demonstrated the presence of costs in switching between the two tasks as discussed by (Rogers & Monsell, 1995). I also derived and tested a novel prediction on the interaction of the pattern of switch costs for task sets compatible and incompatible with the default set, and confirmed our predictions for words, but not for pictures.

References

Besner, D., Moroz, S., & O'Malley, S. (2011). On the strength of connections between localist mental modules as a source of frequency-of-occurrence effects. *Psychological Science*, *22*(3), 393–398.

Cattell, J. (1886). The Time it Takes to See and Name Objects. *Mind*, *11*(41), 63–65.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature*.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology General*, *124*(2), 207.

Seymour, P. H. (1979). *Human visual cognition: A study in experimental cognitive psychology*. Collier Macmillan London.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, *47*(3), 631–650.