Project Q: The Use of Incidental Cues in Digital Flashcard Learning

Darragh Fitzpatrick

S4243943

Department of Psychology, University of Groningen

PSB3E-BTHO: Honours Bachelor Thesis

Supervisor: prof. dr. Hedderik van Rijn

July 23, 2023

Abstract

A goal of digital flashcard learning is that learners form associations between the cue and the answer and thereby learn through the mechanisms of cued retrieval. However, cues are often presented with the same contextual features such as verbal and spatial features. These features may become incidentally associated with the desired response. Using such cues to facilitate a response may reduce the need to deeply process cue information. In this study, the aim was to test whether introducing cue variation by means of spatial and verbal manipulations would be associated with increased response times and knowledge of cue information. Forty-six students participated in a two-block mixed design. Each block consisted of a fifteen-minute digital flashcard learning task followed by a test assessing knowledge of cue information. Results indicated that variations in verbal properties of the cue were associated with increased response time in the learning task. However, there were no differences in response time when cues were manipulated spatially, nor was there any observed benefit of spatial or verbal variability in increasing knowledge of cue information. Despite this, the use of predictive incidentally associated cues was reported by participants. Recommendations are made for further research on the use of incidentally associated cues in digital flashcard tasks.

Keywords: Flashcard Learning, Gaming the system. Incidental Cueing, Contextual Cueing, Underspecification

Project Q: The Use of Incidental Cues in Digital Flashcard Learning

Cue-based retrieval is a learning practice whereby a cue (e.g., a starchy vegetable) serves as an aid to retrieve a target response (e.g., potato) (Moult., 2011). It has been demonstrated that cue-based retrieval is an effective learning practice and as a result, it has been utilized as a task in E-learning environments such as digital flashcard learning (Van Rijn et al., 2009). However, when cues are presented in such tasks, they may occur in tandem with other features that can be just as predictive of the required response. For example, when learning definitions, users may be able to identify the appropriate response merely by the size of the block of text or by a keyword within the definition that discriminates it from alternative cues. In this case, learners would not need to understand or even read the definition to produce the correct response but merely identify features of the cue that are predictive of the answer. Using these incidentally predictive cues may increase the efficiency of a response as identifying visual features or keywords of a cue would likely be quicker than reading an entire sentence. However, the use of these strategies could result in limited comprehension of the intended cues, leading to poorer learning outcomes. This paper seeks to address whether this concern is valid and presents potential solutions to this issue.

Gaming the System

Prior evidence suggests that non-task behaviours are used to achieve goals and avoid engagement with tasks in digital learning systems. Baker et al., (2004) describe a phenomenon in digital tutoring platforms which they term “Gaming the system”. This describes the behaviour employed by learners in digital tutoring platforms that use the design of the learning environment to fulfill the goals assigned. In their first paper describing the phenomenon in a problem-solving task, they focus on behaviours such as help-seeking and inputting answers repeatedly until the correct answer is identified (Baker et al., 2004). They distinguish this from cheating as these behaviours simply take advantage of gaps in the system rather than violating the rules of the system. These behaviours were significantly related to post-test scores, with students who gamed the system performing worse than those who didn't although there were baseline differences in performance already present.

Gaming the system was also unrelated to other off-task behaviours which the authors suggest reflects that those who game the system do so as their goal is performing rather than learning (Baker, 2004). However subsequent research found that performance goals were unrelated to gaming the system behaviours (Baker, 2005). Rather, performance goals were associated with slower responses. As for how gaming-the-system behaviours arise, the researchers suggest that these learners construct knowledge in a "task-specific way" that is, knowledge useful for completing goals within the digital learning task. However, this task-specific learning may be unrelated to learning the material and therefore detrimental to educational development (Baker et al., 2004). It is further argued that gaming the system behaviours may occur particularly in situations that the learner finds difficult. If this is the case, the learner would avoid task engagement in situations where they stand to benefit the most from effort in terms of educational development. Although Baker identified gaming the system behaviours in problem-solving tasks, previous research has not identified if similar behaviours occur in cued-recall learning tasks such as flashcard tasks. Therefore, it is useful to consider the nature of digital flashcard tasks and identify the features of such tasks from which these behaviours could arise.

Visual Context and Contextual Cueing

When a retrieval cue is presented visually, the cue may have the same visual characteristics each time it is presented (e.g., the cue is always presented as a large block of text, or the same pattern of spaces occurs in each presentation). These perceptual regularities may be encoded and become associated with the target response. If the use of these regularities is all that is necessary to generate the target response it would not be necessary to read and encode the information within the cue. Isirada and Isirada (2007) describe the incidental learning of these visual regularities as the learning of the *environmental context* in which learning takes place. The association of environmental context with an item can produce context effects where item retrieval is improved when the context present at encoding is reinstated. In a free-recall task, Isirada and Isirada (2007) found that context effects were present when two different background colours were used in the learning stage. However, no context effects were found when the background colour remained constant for all items. This suggests that having as little as two alternative environmental contexts in a learning task is enough to produce contextual-dependent memory effects (Isirada et al., 2007).

A similar phenomenon is observed in visual search tasks and is referred to as *contextual cueing* (Chun and Jiang, 1998). Contextual cueing occurs when the *visual context* of a stimulus aids in the deployment of attention to particular features or objects within a search task. Contextual cueing occurs outside of conscious awareness and can be described as the incidental learning of repeated visual contexts, (Reber, 1989). Chun and Jiang describe contextual cueing in terms of “memory-based automaticity” (Chun and Jiang, 1998, p.31). This can be understood as a form of incidental learning of repeated visual contexts. The learning of visual contexts appears to be implicit, however, that does not necessarily preclude the learner from becoming aware of these regularities and using them for goal-oriented purposes. Furthermore, it should be noted, that although there is evidence demonstrating the implicit learning of repeated visual contexts facilitating retrieval in free recall and visual search tasks, there is little research on whether this occurs in cued-recall tasks such as in flash-card learning. Within a flash-card learning task, the presentation of a cue can be understood to have a repeated visual context. In this case, the properties such as the size of the text block and length of lines would be the visual context of the cue. The repetition of these contexts could conceivably become reliable predictors and associated with the target response. Learners within flashcard-learning tasks could therefore learn to depend on the visual features of a cue to retrieve a response.

Good Enough Processing and Underspecification

Although the intention behind flash-card learning tasks may be that the user attentively reads and deeply processes the cue in order to remember the information, this may not be necessary to fulfill task goals. Shallow processing of verbal cues may be all that is required to provide a target response in such tasks. *Good-enough processing* of linguistic representations was suggested by Ferreira and colleagues (2002). They proposed that language comprehension did not always prioritize completely accurate representations of information.

Underspecification is a term that describes a situation where a representation is not fully processed to a complete level (Swets et al., 2008). Underspecification and prioritization of information within a sentence are influenced by the task demands at hand (Logačev & Vasishth, 2016; Christianson, 2016). Swets et al., (2008) demonstrated quicker reading times of ambiguous sentences when participants expected superficial comprehension questions rather than more advanced comprehension questions. Further research has also identified how time pressure and reading medium affect the depth of processing and comprehension (Delgado & Salmaron, 2021). Readers who were given online reading material have been found to have weaker comprehension of the material than those given print. If complete comprehension is not necessary to fulfill the task goals, underspecification of the sentence is more likely to occur (Fitzsimmons & Drieghe, 2013).

In a flashcard learning task such as in definition learning, a cue may be the definition of a word for example “a nonhuman primate mammal with the exception usually of the lemurs and tarsiers”, may be the definition of the target word “monkey” (Merriam-Webster, n.d.; Moult., 2011). In this instance, the goal of the task is to respond by inputting the response “monkey”. In order to fulfill this task goal a user may only need to identify a keyword that is not present in any other item within the learning set, for example, the word “primate”. In this case, complete comprehension of the cue is not required and the task demands are low. This raises the question of whether low-task demands may lead to underspecification and therefore poorer comprehension of cue information within a digital flashcard task.

Slimstampen and the ACT-R

Slimstampen is a digital learning platform that implements cue-based retrieval in the form of a flashcard-style learning task. In these tasks, a cue is presented visually and learners are instructed to type in the corresponding answer. Slimstampen is an adaptive learning system; a system that may use various methods such as machine learning to enhance performance and provide a more personalized learning experience (Weber, 2012). As part of this adaptive learning system, aspects of the cognitive architecture ACT-R are employed in Slimstampen’s algorithm. In particular, the system makes use of models within the declarative memory module of the ACT-R to predict learners' future performance and assess knowledge of learned items within the flashcard learning task (Van Rijn, 2009; Anderson et al., 2014).

Within the declarative memory module of the ACT-R, the probability of retrieving an item or *chunk* depends on its level of activation (Anderson et al., 2014). The activation level of a chunk is determined by its recency and frequency of activation. Each chunk also has an activation threshold; a minimum level of activation required for the chunk to be successfully retrieved, and a rate of decay in activation. Response times and accuracy in ACT-R reflect the activation of a chunk and therefore the ease of retrieval. This allows for estimates of how well an item is learned.

The declarative memory module of ACT-R can also be used to explain and model memory effects such as the *spacing effect.* The spacing effect, first identified by Ebbinghaus, refers to the observation that the optimal time to learn items is when these items are temporally separated (e.g., learning is more effective in an AB AB presentation order than AA BB order) (Ebbinghaus, 1885). A model proposed by Pavlik and Anderson (2005) accounts for this effect within the ACT-R framework.

The model of spacing effects by Pavlik and Anderson (2005) has been implemented into Slimstampen and combined with the testing effect to schedule the presentation of items within a learning session. The testing effect refers to the finding that learning outcomes are better for those who are tested on items they learned and successfully retrieve, than if they learned the same items and had not been tested at all (Roediger, 2006). By balancing both of these effects and incorporating them into Slimstampen, the system attempts to optimize the schedule of item presentation within a learning session. User metrics such as accuracy and response times provide can be used to determine the activation of an item, this can be compared with the model's prediction to determine how items should be spaced. This adapts the user's learning experience so they received an optimized schedule of testing on items. Slimstampens adaptive learning system has been shown to be more effective than simple flashcard learning (Van Rijn et al., 2009). However, the learning environment may still remain prone to learners using *incidental cueing* strategies to provide a response, which could undermine the learning outcomes of users.

Present Study

It is not understood the effect that incidental learning of superficial cue features has on the learning process in digital flashcard learning tasks. Potential evasive behaviours in such tasks may include the use of visual features of a cue; and or inference based upon incomplete verbal processing of the cue sentence. If such behaviours are indeed present in the learning session, within-task performance may not be reflective of learning the intended information but only superficial contingencies that are predictive of the target response. Such behaviours may therefore undermine the effectiveness of e-learning systems.

This paper seeks to address the effect that spatial and linguistic variability of cues has on the retention of cue information and response times in a flashcard learning session within Slimstampen. By increasing the number of varying cues that correspond to a given to-be-learned word, we hope to decrease the predictive validity of incidental cues. We expect increased processing and retention of cue information. In particular, we will adjust the spatial layout of the cue to change the visual characteristics and replace words in cues with synonyms to alter the verbal characteristics. We propose that introducing cue variability along these dimensions will require the learner to devote more attention towards the cue to respond with the correct answer, which we expect to be reflected by increased response times within the learning session. In turn, we expect that learners will be able to retrieve more information about the learned concepts when provided with what was previously the target response. This will provide insight into how well cue information is processed and retained across contexts. To address these predictions, the following hypotheses will be tested.

Hypothesis 1: Cue variation will be associated with increased response times in a Slimstampen learning session

A: Variation in verbal properties of a cue increases will be associated with longer response times within flashcard learning session

B: Variation in spatial properties of a cue increases will be associated with longer response times within flashcard learning session

Hypothesis 2: Cue variation will be associated with increased performance on a test of cue information:

A: Those with variation in verbal properties of a cue will perform better than those with no variation

B: Those with variation in spatial properties of a cue will perform better than those with no variation

Method

Participants

Undergraduate psychology students from the University of Groningen were recruited for participation in this study. Most enrolled for participation in the study through the faculty research portal SONA. Others were acquaintances of the researchers and voluntarily participated in the research. After confounding data such as those who did not complete the learning session(s) were removed a final sample of forty-six remained. Although exact age was not measured, being younger than 18 or older than 25 was an exclusion criterion, and no further data was collected for these participants.Individuals'sfirst languagee was not recorded although participants were informed the study would be conducted in English. Participants provided informed consent at the beginning of the study. The informed consent form followed the guidelines of and was approved by the ethics committee at the University of Groningen. The study code for this research was (PSY-2223-S-0355).

Stimuli

This study consisted of two blocks. The learning set in both blocks consisted of 25 definitions taken from a cognitive psychology book "*Cognition: Exploring the Science of the Mind*" (Reisberg, 2018). The definitions differed between the first and second block. In the first block all participants were presented the same definitions. In the second block the definitions were presented in one of three ways *default, spatial,* and *synonym*.

Default, was a simple typed version of the definition with no alterations made, participants in the default condition saw only this version of the cue. In the two experimental conditions items were presented in two ways, the default version, and the respective manipulated version. To make this clearer, for each item in the second block there were three variations, A,B,and C. In the second learning session, participants were assigned to one of three conditions, A only, A and B, or A and C. In the spatial adapted definition, all linguistic elements of the original cue remained, however, one line break was inserted at an arbitrary location within a sentence of the cue, thus altering the visual appearance of the cue. The line break was inserted within a sentence rather than at the beginning/end. In the synonym condition, the words from the default condition were replaced with synonyms. A balance was sought between retaining the semantic integrity of the original word and minimizing the visual differences elicited from the changes in particular words.

Procedure

The study was conducted using Qualtrics and contained two blocks. Both blocks consisted of a learning phase followed by a testing phase. In the first block, each participant was given a link to a fifteen-minute Slimstampen session, in which they were presented the first set of definitions. Participants were instructed to treat this session as they typically would while preparing for an exam. When first presented with an item, participants did not receive an input option but were expected to memorize the cue-answer pair. All presentations after this required the participant to respond to the cue with the corresponding answer. The order in which new items were introduced was random. The presentation order of items after initial presentation was determined by the Slimstampen algorithm.

If a participant provided the correct term, they were provided with feedback indicating a “correct” response, lasting 0.4 seconds. If they provided an incorrect response they were provided with an “incorrect” feedback response and provided with the correct response, for 0.6 seconds. After the fifteen-minute learning session had ended, participants were instructed to return the Qualtrics survey and were subsequently provided with a link to an online Tetris distractor task which lasted for two minutes. After this, participants were directed to the test. Participants were presented a Qualtrics page providing them with all of the terms they had learned in the previous Slimstampen session with text entry boxes below each term. These terms were previously the answers they were instructed to respond with within the Slimstampen session. The participants were asked to fill in as many details as they could remember about the corresponding definition.

The same procedure was carried out for the second block as the first except in this block participants were randomly assigned to one of the three potential conditions, default, spatial, or synonym. Besides this change, the two blocks were otherwise identical. After completion of the second post-test, participants were provided two questions that served as validity checks and sources of qualitative data. The first question concerned whether they changed tactics between the first and second learning blocks. The second question asked participants what they expected the hypothesis of the research to be.

Response Coding

A grading rubric was created for the scoring of responses. This grading rubric was formed to not require a verbatim definition but rather to assess whether key terms and concepts were understood. The grading rubric for all items can be found in Appendix A. Participants' responses were given scores according to how many of the correct answer requirements were fulfilled.

Results

Hypothesis 1: Learning Session Response Times

The first hypothesis suggested that response times in the adapted cue presentation conditions would be greater than those in the default presentation condition. To assess this, an analysis of covariance (ANCOVA) was conducted with median response times in milliseconds on the second learning block as the dependent variable with median response time on the first block as a covariate. A decision was taken to use the median response time differences as some outliers had response times exceeding one minute. The median response times differed significantly across the three conditions spatial (M_{adj} = 4463.3), synonym (M_{adj} = 5240.2), and spatial (M_{adj}=4651.2). An effect of condition on response time difference was found at the $p<.05$ level F (2,43) = 4.198, p= 0.022. These results fall in line with the hypothesis that increasing cue variation would lead to longer response times within the Slimstampen learning session. However, simple contrasts revealed that only the synonym condition t(42), p=.013, r=0.37, differed significantly from the default condition thus supporting hypothesis 1A but not hypothesis 1B.

Figure 1

Adjusted Means for Response Time in Second Learning Block

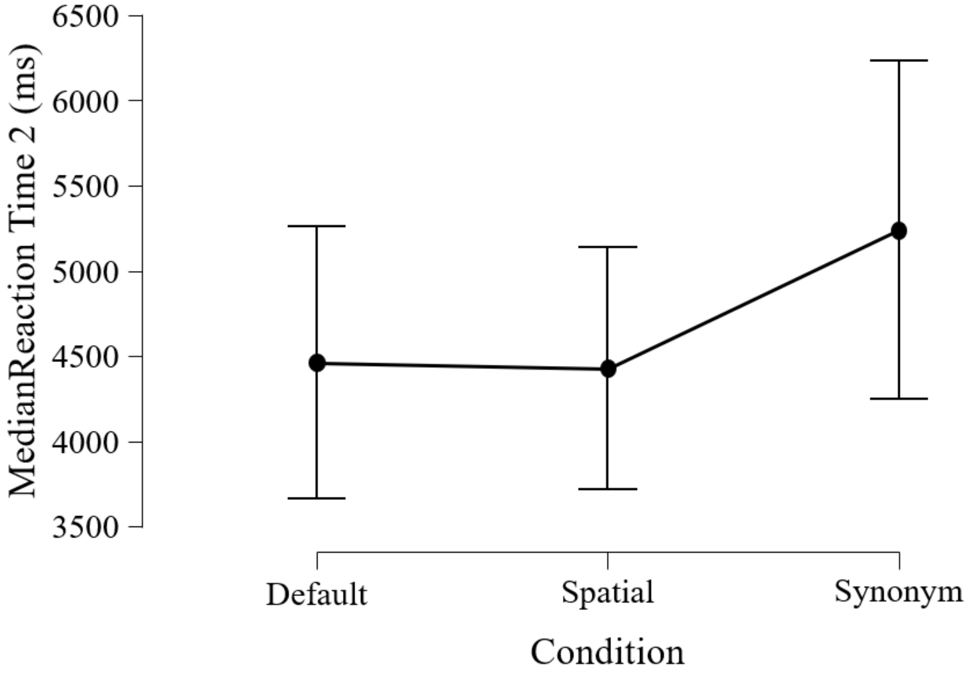


Table 1

Simple Contrasts of Median Response Time in Second Learning Session

Comparison	Estimate	SE	df	t	p
Spatial - Default	161.060	423.483	42	0.380	0.706
Synonym - Default	1136.273	436.517	42	2.603	0.013

Hypothesis 2: Cue Information Retention

The second hypothesis concerned whether increasing the variance of the cue presentation in particular by manipulating the spatial presentation of the cue, and the words used in the cue would be associated with greater knowledge of cue information. Accuracy was determined as the number of correct divided by the number of unique items a participant was presented in the Slimstampen learning session. To test this hypothesis an ANCOVA was conducted to assess performance between group differences in accuracy in the second post-test. Adjusted means for post-test accuracy were similar across the Spatial (M_{adj}=0.48), Synonym (M_{adj}=0.543), and Default (M_{adj}=6.01) conditions. The Default participants slightly outperformed both the Synonym and Spatial participants. An ANCOVA revealed that there was no significant effect of condition on the performance of the second post-test at the $p<.05$ level, F (2,43) = .448 p=0.642. Contrasts also revealed neither the Spatial t(42), p=0.571, r=.087, nor the Synonym t(42), p=0.35, r=0.14, participants differed significantly from Default. This is in contrast to both hypothesis 2A and hypothesis 2B. These results are contrary to the expectation that increasing cue variation would be associated with greater knowledge of cue information in the post-test.

Figure 2

Adjusted Means of *Accuracy in Second Post-Test Across Condition*

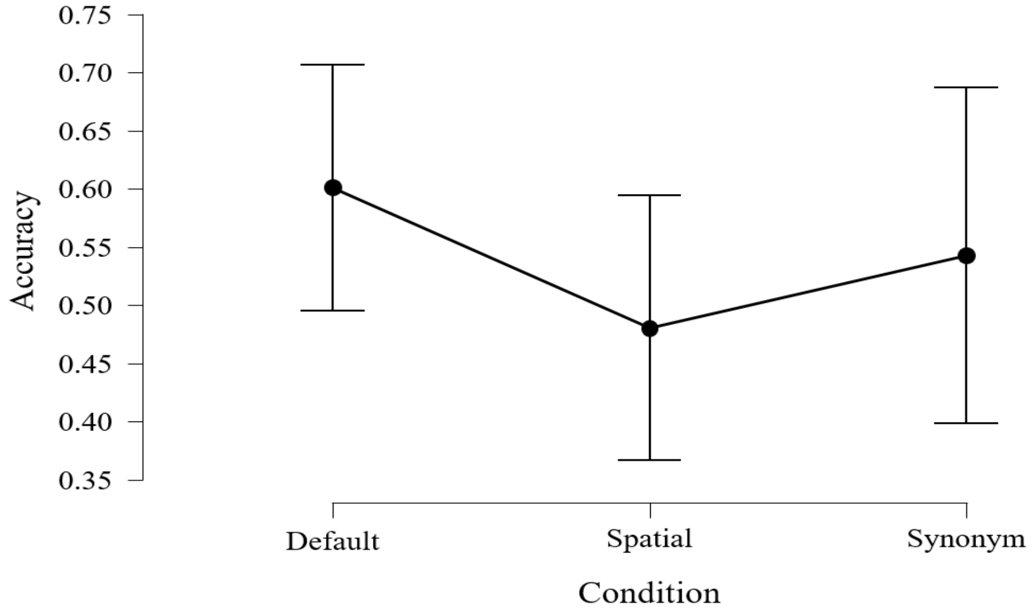


Table 2

Simple Contrasts for Second Post-test Accuracy Across Condition

Comparison	Estimate	SE	df	t	p
Spatial - Default	0.042	0.074	42	0.571	0.571
Synonym - Default	0.066	0.071	42	0.929	0.358

Additional Analysis

Additional data analyses were carried out in order to aid in the interpretation of the data and provide further insight into participants behaviour during the Slimstampen learning sessions.

High and Low Performers

Within-group and between-group median, splits were conducted on the first post-test performance to establish baseline differences in the sample. Within group median split revealed a somewhat even split within each group. However, a median split of performance across groups revealed a significant difference in those classified as high and low performers. In particular of those in the default condition, 86% were classified as high performers in contrast to 37.5% and 46% in the spatial and synonym conditions respectively.

Table 3

High and Low Performers Across Condition

	Condition			
GenH/L	**Default**	**Spatial**	**Synonym**	**Total**
High	13	6	7	26
Low	2	11	8	21
Total	15	17	15	47

Table 4

Chi-Squared Test of High and Low Performers Across Condition

	Value	**df**	**p**
X^2	9.176	2	0.010
N	47		

Reactive Measure Assessment

As counterbalancing was not included in the study design it is possible that participants employed different learning strategies in the second learning session in order to improve performance and thus decreasing the internal validity of the study. In a post-study questionnaire, participants were given the following prompt with a written response option "Did you change learning strategy between the first and second learning session? If so, explain how your strategy changed". After data collection, these responses were coded as Change/No Change, as well as coded into subsets concerning the nature of the strategy change; *verbatim, meaning, reading,* and *keywords*. In total 69.5% of participants indicated that they had changed their learning strategy in the second Slimstampen learning session. Twenty-four participants reported trying to remember the meaning of the word they were learning, seven reported trying to memorize the definition verbatim. Two participants indicated they read the cue in full every time and two other participants indicated they used keywords in order to produce their answers rather than reading the entire text.

Table 5

Change in Tactic Across Condition

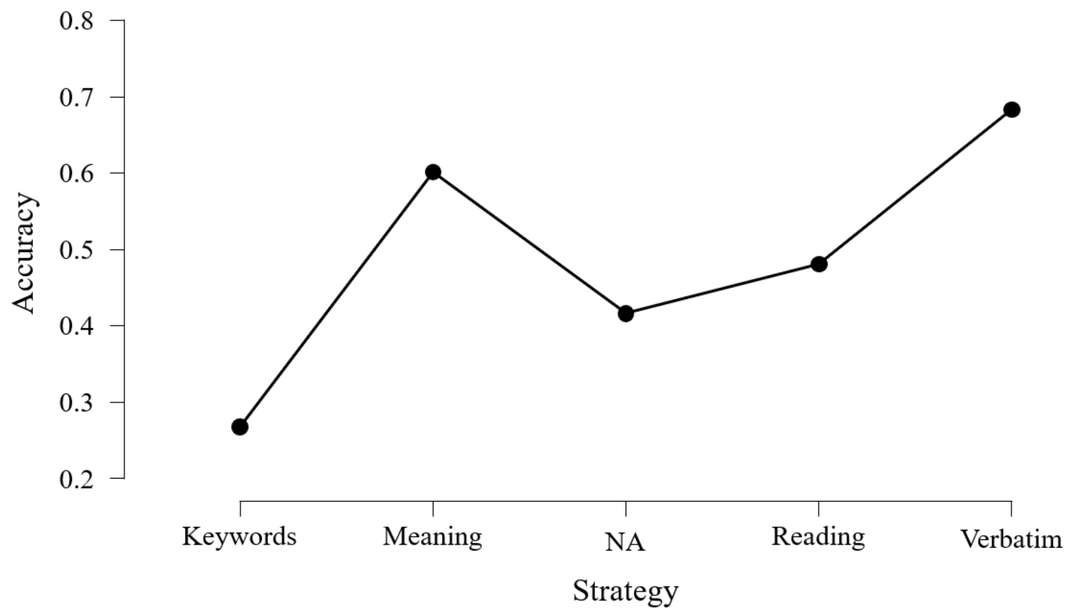
	Condition			
V204	**Default**	**Spatial**	**Synonym**	**Total**
Keywords	1	1	0	2
Meaning	7	4	11	22
NA	4	6	4	14
Reading	1	1	0	2
Verbatim	2	5	0	7
Total	15	17	15	47

ANCOVA Including Exploratory Variables

An ANCOVA was conducted with accuracy in the second post-test as dependent variable and including change in tactic strategy as a fixed factor and accuracy in the first post-test as covariate. There were significant differences in accuracy across the different tactics used in the second learning session at the $p<.05$ level, $F(4,41) = 5.949$, $p < .01$. In particular, posthoc Tukey comparisons revealed that those who used keywords performed significantly worse than those who used a meaning comprehension tactic $M=-.382$, $p<.01$, and those who tried to learn the definition verbatim $M=-487$, $p<.01$.

Figure 3

Adjusted Means of Post-Test Accuracy Across Learning Strategies



Self-Report Use of Incidental Cues

Participants frequently reported the use of self-generated cues within the Slimstampen learning sessions. This was reported both in the open question and directly to the researchers after completing the study. These incidental cues as expected were most frequently keywords and spatial features respectively.

Discussion

In this study, we sought to assess whether cue variability would be associated with a greater knowledge of cue information and longer response times in a flashcard-style cue-based retrieval task. It was expected that increasing cue variability, both spatially and verbally would reduce the learner's ability to use incidentally associated cue-response contingencies to retrieve responses within the learning session. The hypotheses were that cue variability would be associated with longer response times within the learning session, and a deeper comprehension of the material, as assessed by retrieval of cue information in the second post-test. The results indicate that including multiple cue variants was indeed associated with increased response time within the Slimstampen learning session. However, contrasts revealed this was only statistically significant when cues varied verbally. What became evident from the data however was that cue variability in spatial or verbal features did not aid in retaining conceptual information of the cue.

Findings and Implications

In line with our first hypothesis, the data obtained from the second Slimstampen learning session indicated slower response times according to condition. It was revealed that the synonym condition had the longest median response times among all groups. This proved to be significantly different from the default condition, in line with hypothesis 1A. The decreased predictive validity of salient and highly efficient cues would conceivably require the learner to dedicate more cognitive resources to determine the relevant response. That said, differences in response time may reflect increased reading time, retrieval time, or a range of other cognitive processes. Therefore, from response time data alone, it cannot be concluded whether introducing cue variation within the learning session led participants to pay more attention to the contents of the cue or whether it simply made retrieval more difficult.

It must also be noted that the removal of the covarying incidental cues required the use of multiple alternatively structured cues, which in and of themselves could result in an increased response time as would be hypothesized under the fan effect (Anderson, 1974). The fan effect describes the finding that more associations with a concept will result in a weaker associative strength between the associated nodes and the concept. Hence, the associative strength of a definition with one cue will be greater than multiple cues. Multiple cue-target associations would in this case evoke a longer response time in retrieving the target (Radvansky, 1991). With this consideration, it should be noted that the increased response time in the synonym condition may reflect a general decrease in retrieval efficiency as multiple cue-target associations must be learned.

Although under the fan-effect multiple cues may introduce a general reduction in response time (Thomson, 2017), contrary to hypothesis 1B contrasts revealed no significant difference between the median response times of the spatial and default condition. This may suggest that more than just the number of alternative cues but also the means by which the cued differed played a role in increasing response time. A possible explanation for this could be that the spatial differences between cues were not strong enough to influence retrieval time. The synonym condition on the other hand may have been a more extreme manipulation and therefore interfered with retrieval processes to a greater extent. This falls in line with prior research on the use of contextual cues in Stroop paradigms by Crump et al., (2007). In their study, shape-based and location-based contexts were used and stimulus congruency was more likely to occur with particular shapes and locations. Congruency effects were identified for the location condition, but the researchers found that no shape-based Stroop congruency effect was present. It is important to note that awareness of the contingencies was not present in either condition. This demonstrates that some visuospatial characteristics are more effective at producing a context effect than others. This could be the underlying cause for a lack of differences in response time between the spatial and default condition.

With regard to post-test accuracy in the second block, there were no significant differences between groups. This was contrary to both hypothesis 2A and hypothesis 2B. The lack of differences across groups revealed no apparent benefit of the cue variation on being able to retrieve cue information in the post-test. One possible explanation for this is due to the effect of transfer-appropriate processing (Morris et al. 1977). Transfer-appropriate processing describes the finding that performance is improved if the mental processes engaged during testing match those that were engaged during encoding (Yanes et al., 2019). Blaxton (1989) provided evidence that the dissociation in performance between differing memory tasks can be considered a result of whether they share the same mental performance operations. In this study to evaluate conceptual knowledge of cue information, participants were provided the previous target word and asked for the cue associated with it. However, this by necessity switched the task between encoding and retrieval perhaps leading to more difficult-to-retrieve memory traces. It is therefore unclear how much the change between learning and retrieval task contexts may have affected the performance on the post-test.

The exploratory analyses provided some insights into the learning within Slimstampen and notably, the strategies employed by participants within the learning session. As expected, those who relied on keywords performed the poorest in the post-test. This provides further evidence that the shallow processing of cues is associated with weaker comprehension and retention of cue information (Swets et al., 2008; Delgado & Salmaron, 2021). These results also build on prior gaming the system research and suggest that gaming the system behaviours may also include the use of incidentally predictive cues within a flashcard learning system. These results should also be taken with a grain of salt however as these reported tactics were specifically related to a change in tactic. Those who reported no change and were placed within the NA group may well have employed similar tactics in both sessions.

In summary, findings support only one hypothesis, that increased cue variability in the synonym condition would be associated with longer learning session response times. However, mixed results preclude any clear explanation for this finding. Nevertheless, it should be noted that additional data revealed evidence for the presence of non-task-relevant behaviours in Slimstampen. Unprompted, participants informally reported in both the post-test questionnaire and directly to the researchers, that they routinely used such incidental cues in both the study and whilst using Slimstampen in an academic context. While this was self-reported it is apparent that this is a strategy that some learners employ within Slimstampen. Although these tactics may be used within Slimstampen or other digital flashcard learning systems, the manipulation of adding cue variability does not appear to be effective in increasing knowledge of cue information. In terms of implications for educators utilizing digital flashcard systems, students should be encouraged to not rely on the use of incidental cues and reminded that the use of such cues will undermine their own academic development.

Limitations and Future Directions

The results of this study may be partially a result of methodological limitations. First and foremost, as previously mentioned, the cued retrieval task switched the order of the cue and the target response. The reversal of cue and target may have introduced increased difficulty of the task and learners may have relied on differing cognitive mechanisms than in the learning task (Blaxton,1989). This decision was still in line with the research question. The intention was to address the degree of conceptual knowledge and address how well the cue; in this case the definition, was associated with the given target word response. This was considered a more realistic situation of testing and perhaps a more appropriate measure of assessing the effectiveness of digital learning systems in an educational context. However, an alternative and less difficult choice of test task may reveal more about the level of processing and cue knowledge.

One possible alternative task may be a multiple-choice recognition task between various cues with similar perceptual features yet slightly different meanings. For example, “The book on the table was read” with an alternative choice of “The book on the table was red”. Additional analyses also indicated that there was a large degree of learning tactic change in the second learning session. The reactivity of our assessment therefore may have introduced confounding behavioural changes in participants. This is problematic as there is a possibility that the data may be a result of a change in learning tactics rather than the intended experimental manipulation. A simple between-participant design with one learning phase and one testing phase may reduce the occurrence of these tactic changes.

Within the design of the learning task, there are potential limitations in the selection of items used. In hindsight, the selection of to-be-learned items shared a high degree of conceptual similarity. For instance, there were multiple items referring to distance cues (see Appendix A). It can be assumed that if a distance cue is a new concept to a learner, it may require extra cognitive effort to differentiate the different subsets of this phenomenon. Task difficulty could be reduced by developing a standardized set of to-be-learned items that are to be used in such experiments. This would help resolve potential issues such as conceptual overlap or excessive difficulty in learning items.

Given the self-reported use of incidental cues in this study, there is now empirical qualitative evidence for their use in digital flashcard learning tasks. A potential avenue for future research is to gather more comprehensive qualitative data from learners who use these systems. Self-report measures may help to clarify the degree to which various incidental cues are used and the contexts in which they are employed. For example, it may be insightful to see whether these behaviours are reported more frequently in low performing learners as was found by Baker et al., (2004). Furthermore, qualitative data may help provide clarity on the role of motivational factors in gaming-the-system behaviours in flashcard learning tasks. Motivation could prove to be a reliable predictor of whether learners use such behaviours as previous research suggests that self-efficacy is a reliable determinant of whether learners cheat (Nora & Zhang, 2010). Besides revealing the contexts in which the use of incidental cues occurs, expanding understanding of learner experience may also help explain the quantitative findings of this study. This includes the slowed response times present in the synonym condition. Research from vigilance tasks that require sustained attention routinely shows a “vigilance decrement”, a slowing in response time across the vigilance task. Factors such as task difficulty have been demonstrated to increase this vigilance decrement (Helton et al., 1986). A long flashcard learning session similarly requires sustained attention, it may therefore be interesting to enquire as to whether a vigilance decrement contributed to the longer response times seen within the synonym condition.

Finally, efforts should be made to more concretely establish the phenomena described in this paper within existing theoretical frameworks such as contextual cueing, gaming-the-system, and good enough processing research. For this purpose, it should not be understated the potential benefit of including other measures during the learning session. In particular, the use of eye-tracking may help to understand the biasing of attention and explain the mechanisms underlying the use of visual incidental cues and keywords as was reported by participants in this study. Eye-tracking has been utilized in contextual cueing paradigms. For instance, research by Liechty et al., (2011) demonstrated fewer saccades and lower saccade amplitudes in a contextual cueing task, which indicates efficient search for the target. Thus eye-tracking can be used to assess attention toward features of a contextual cue. It remains to be seen whether the same occur eye-movement behaviours occur in a flashcard learning context. Eye-tracking has also been used to demonstrate the word-skipping of predictable words within underspecification research (Fitzsimmons et al., 2013). Again, it may be revealing to see whether the same word-skipping phenomenon occurs when keywords are available in a flashcard cue in comparison to when they are not. Therefore, the use of eye-tracking in future experiments concerning incidental cueing in flashcard tasks can be used to establish a concrete connection to existing theoretical frameworks.

Conclusion

In conclusion, there appears to be evidence that the use of incidentally predictive features of cues is present within the flashcard learning task in Slimstampen. There is partial evidence that increasing variability of these cues by means of increasing cue variation linguistically interferes with response times within the learning session. However, data indicated that this led to no change in knowledge of cue information. Future research should seek to improve the procedure for assessing the use of such cues. In order to more directly assess the impact of multiple distinct cues on attention within the learning session methods such as eye-tracking and qualitative data analysis may be used.

References

Anderson, J. R. (1976). *Language, Memory, and Thought*. http://ci.nii.ac.jp/ncid/BA00681819

Anderson, J. R., & Lebiere, C. J. (2014). *The atomic components of thought*. Psychology Press.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting Student Misuse of Intelligent Tutoring Systems. In *Springer eBooks* (pp. 531–540). https://doi.org/10.1007/978-3-540-30139-4_50

Baker, R. S., Roll, I., Corbett, A. T., & Koedinger, K. R. (2005). Do Performance Goals Lead Students to Game the System. In *Artificial Intelligence in Education* (pp. 57–64). http://www.columbia.edu/~rsb2162/BRCKAIED2005Final.pdf

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*(4), 657–668. https://doi.org/10.1037/0278-7393.15.4.657

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603

Chun, M. M., & Jiang, Y. V. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Crump, M. J. C., Vaquero, J. M. M., & Milliken, B. (2008). Context-specific learning and control: the roles of awareness, task relevance, and relative salience. *Consciousness and Cognition*, *17*(1), 22–36. https://doi.org/10.1016/j.concog.2007.01.004

Delgado, P., & Salmerón, L. (2021). The inattentive on-screen reading: Reading medium affects attention and reading comprehension under time pressure. *Learning and Instruction*, *71*, 101396. https://doi.org/10.1016/j.learninstruc.2020.101396

Ebbinghaus, H. (2013). Memory: a contribution to experimental psychology. *PubMed*. https://doi.org/10.5214/ans.0972.7531.200408

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Fitzsimmons, G., & Drieghe, D. (2013). How fast can predictability influence word skipping during reading? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *39*(4), 1054–1063. https://doi.org/10.1037/a0030909 https://doi.org/10.5167/uzh-6226

Helton, W. S., & Warm, J. S. (2008). Signal salience and the mindlessness theory of vigilance. *Acta Psychologica*, *129*(1), 18–25. https://doi.org/10.1016/j.actpsy.2008.04.002

Isarida, T., & Isarida, T. (2007). Environmental context effects of background color in free recall. *Memory & Cognition*, *35*(7), 1620–1629. https://doi.org/10.3758/bf03193496

Liechty, M., & Madhavan, P. (2011). An eye movement analysis of contextual cueing effects. *Proceedings of the Human Factors and Ergonomics Society . . . Annual Meeting*, *55*(1), 222–226. https://doi.org/10.1177/1071181311551046

Logačev, P., & Vasishth, S. (2016). Understanding underspecification: A comparison of two computational implementations. *Quarterly Journal of Experimental Psychology*, *69*(5), 996–1012. https://doi.org/10.1080/17470218.2015.1134602

Merriam-Webster. (n.d.). Monkey. In Merriam-Webster.com dictionary. Retrieved July 23, 2023, from https://www.merriam-webster.com/dictionary/monkey

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/s0022-5371(77)80016-9

Moult, M. (2011). Cued recall. In *Springer eBooks* (pp. 751–752). https://doi.org/10.1007/978-0-387-79948-3_1116

Nora, W.L.Y., Zhang, K.C. Motives of cheating among secondary students: the role of self-efficacy and peer influence. *Asia Pacific Educ. Rev.* **11**, 573–584 (2010). https://doi.org/10.1007/s12564-010-9104-2

Pavlik, P. I., Jr, & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: an activation-based model of the spacing effect. *Cognitive science*, *29*(4), 559–586. https://doi.org/10.1207/s15516709cog0000_14

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 940–953. https://doi.org/10.1037/0278-7393.17.5.940

Reber, A. C. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219

Reisberg, D. (2018). Cognition: Exploring the science of the mind. W. W. Norton & Company.

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.xSwets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. https://doi.org/10.3758/mc.36.1.201

Thomson, R. R., Harrison, A., Trafton, J. G., & Hiatt, L. M. (2017). An Account of Interference in Associative Memory: Learning the Fan Effect. *Topics in Cognitive Science*, *9*(1), 69–82. https://doi.org/10.1111/tops.12244

van Rijn, D., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving Learning Gains by Balancing Spacing and Testing Effects. In A. Hoses, D. Peebles, & R. Cooper (Eds.), *Proceedings of the 9th International Conference on Cognitive Modeling* (pp. 108-114). [187].

Weber, G. (2012). Adaptive learning systems. In *Springer eBooks* (pp. 113–115). https://doi.org/10.1007/978-1-4419-1428-6_534

Yanes, D., Frith, E., & Loprinzi, P. D. (2019). Memory-Related Encoding-Specificity Paradigm: Experimental Application to the Exercise Domain. *Europe's journal of psychology*, *15*(3), 447–458. https://doi.org/10.5964/ejop.v15i3.1767

APPENDIX A.

Default Definition and Grading Rubrice – Learning Session 1				
Visual Features: The constituents of a visual pattern - vertical lines, curves, diagonals, and so on - that, together, form the overall pattern. **Grading**: correct if a) mention that they are the parts of a visual pattern/stimulus b) give an example	**Viewpoint-Independent Recognition** A process in which the ease or success of recognition does not depend on the perceiver's particular viewing angle or distance with regard to the target object. **Grading:** correct if a) mention independence of object recognition from perspective to target	**Geon** One of the basic shapes proposed as the building blocks of all complex three-dimensional forms. **Grading**: correct if a) mention that they are shapes/blocks b) mention that they make up all complex forms/objects	**Motion Parallax** A distance cue based on the fact that as an observer moves, the retinal images of nearby objects move more rapidly than of objects farther away. **Grading**: correct if a) indicate it is a distance cue b) describe that it is a difference in retinal image/sensation of movement c) mention direction of perceived motion change correctly	**Repetition Priming** A pattern of priming that occurs simply because a stimulus is presented a second time; processing is more efficient on the second presentation. **Grading**: correct if a) mention that it is based on repeating presentations b) mention that processing gets faster/more efficient
Tachistoscope An old device that allows the presentation of stimuli for precisely controlled amounts of time. **Grading**: correct if a) mention it is used to present stimuli b) mention that interval/speed of presentation can be controlled c) that it is an (old) device	**Distance Cues** Information available that allows the perceiver to judge an object's position. **Grading**: correct if a) mention that it describes cues/info that allow to judge position/depth/distance of object	**Binding Problem** The problem of reuniting the various elements of a scene, given that these elements are initially dealt with by different systems in the brain. **Grading**: correct if a) mention that different elements/aspects are processed differently by brain b) mention that it concerns binding of separate processes	**Local Representation** A representation in which information is encoded in some small number of identifiable nodes. **Grading**: correct if a) mention that encoding/representation of information b) is only in small number of identifiable nodes/not spread across multiple nodes	**Neuroimaging Techniques** Noninvasive methods for examining either the structure or the activation pattern within a living brain. **Grading**: correct if a) mention that it can be used for structure/activation pattern b) mention that they are noninvasive/non-damaging
Word Superiority Effect The data pattern in which research participants are more accurate and more efficient in recognizing words (and wordlike letter strings) than they are in recognizing individual letters. **Grading**: correct if a) mention word recognition over letter recognition b) mention that it improves RT and accuracy	**Recency** Whether a detector or node has been activated in the recent past. **Grading**: correct if a) mention that it concerns detectors/nodes/neurons b) mention that it concerns recent activation	**Parallel Processing** A system in which many steps are going on simultaneously. **Grading**: correct if b) mention that happen in parallel/concurrently/simultaneously	**Conjunction Error** An error in perception usually due to attention's oveload in which a person correctly perceives what features are present but misperceives how the features are joined. **Grading**: correct if a) mention it concerns perception b) mention the role of attention c) mention that it concerns the binding/joining of features	**Optic Flow** The pattern of change in the retinal image in which the image grows larger as the viewer approaches an object and shrinks as the viewer retreats from it. **Grading**: correct if a) mention it concerns changes in retinal image/sensation b) mention direction of changes correctly
Shape Constancy The achievement of perceiving the constant shape of objects despite changes in the shape of the retinal image due to variations in viewing angle. **Grading**: correct if a) mention perception of constant shape b) mention that happen despite changes in retinal image/sensation/input	**Linear Perspective** A cue for distance based on the fact that parallel lines seem to converge as they get farther away from the viewer. **Grading**: correct if a) mention that it is distance/depth cue b) mention that it is based on converging lines	**Cognitive Neuroscience** The scientific study of the brain and the nervous system to understand humans' mental functioning. **Grading**: correct if a) mention it is a study/field and mention it concerns the nervous system and/or brain c) mention that it aims to understand mental functioning/cognition	**Inhibitory Connection** A link from one node, or detector, to another, such that activation of one node decreases the activation level of the other. **Grading**: correct if a) mention that it is a link between nodes/detectors/neurons b) mention that activation of one decreases activation of other	**Interposition** A monocular distance cue that relies on the fact that objects farther away are blocked from view by closer objects that happen to be in the viewer's line of sight. **Grading**: correct if a) mention it is a distance cue b) mention it is based on information about blocked objects

Activation Level A measure of the current status for a node or detector. **Grading**: correct if a) mentions activity/status b) mentions it is for nodes/detectors/neurons	**Perceptual Constancy** The achievement of perceiving the constant properties of objects despite changes in the sensory information due to changes in our viewing circumstances. **Grading**: correct if a) mention that properties/features are perceived the same b) that this occur even when viewing circumstances/conditions change	**Figure/Ground Organization** The processing step in which the perceiver determines which aspects of the stimulus belong to the central object and which aspects belong to the background. **Grading**: correct if a) mention it concerns distinction between background and stimulus b) mention it is a step in processing/perception	**Response Threshold** The quantity of information or activation needed to trigger a response. **Grading**: correct if a) mention that it needs to be met for a response/reaction b) mention that it concerns activation and/or information/input	**Neural Synchrony** A pattern of firing by neurons in which neurons in one brain area fire at the same time as neurons in another area indicating that they respond to the same stimulus. **Grading**: correct if a) mention synchronous firing of neurons b) mentions that they can be in different regions/parts of brain c) mention that it indicates they are responding to same stimulus