THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2015
Vol. 68, No. 3, 543–567, http://dx.doi.org/10.1080/17470218.2014.961933

Routledge
Taylor & Francis Group

Once and for all—How people change strategy to ignore irrelevant information in visual tasks

Robert Gaschler[1,2], Julian N. Marewski[3,4], and Peter A. Frensch[5]

[1]Interdisciplinary Laboratory Image Knowledge Gestaltung, Humboldt-Universität Berlin, Berlin, Germany
[2]Department of Psychology, Universität Koblenz-Landau, Landau, Germany
[3]Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany
[4]Faculty of Business and Economics (HEC), Department of Organizational Behavior, University of Lausanne, Lausanne, Switzerland
[5]Department of Psychology, Humboldt-Universität Berlin, Berlin, Germany

Ignoring irrelevant visual information aids efficient interaction with task environments. We studied how people, after practice, start to ignore the irrelevant aspects of stimuli. For this we focused on how information reduction transfers to rarely practised and novel stimuli. In Experiment 1, we compared competing mathematical models on how people cease to fixate on irrelevant parts of stimuli. Information reduction occurred at the same rate for frequent, infrequent, and novel stimuli. Once acquired with some stimuli, it was applied to all. In Experiment 2, simplification of task processing also occurred in a once-for-all manner when spatial regularities were ruled out so that people could not rely on learning which screen position is irrelevant. Apparently, changes in eye movements were an effect of a once-for-all strategy change rather than a cause of it. Overall, the results suggest that participants incidentally acquired knowledge about regularities in the task material and then decided to voluntarily apply it for efficient task processing. Such decisions should be incorporated into accounts of information reduction and other theories of strategy change in skill acquisition.

Keywords: Information reduction; Strategy change; Shortcut; Skill acquisition.

When shopping online and also when performing many other activities in our daily lives, we tend to ignore information. For instance, when driving a car, we (hopefully) only attend to those pieces of information in the visual array that are relevant to maintaining the car on the intended trajectory. To buy a plane ticket or a book online, often one has to explicitly agree with the transaction conditions imposed by the corresponding online shop. To agree with these conditions, one typically has to click on an icon on the store website. Upon the first visit to an online shop, one may invest some time in actually reading these conditions. During subsequent shopping activities one is likely to just click on the icon, completely ignoring the legal information the shop offers. Yet, the data available on how exactly seemingly irrelevant information is ignored in such cases offer a surprising perspective.

Correspondence should be addressed to Robert Gaschler, Department of Psychology, Universität Koblenz-Landau, Fortstraße 7, D-76829 Landau, Germany. E-mail: gaschler@uni-landau.de

We thank Hilde Haider, Pierre Perruchet, Erik M. Altmann, Richard Carlson, Mark Haselgrove, and William Raymond for helpful suggestions, as well as Anita Todd and Kate Koennecke for help with editing this manuscript.
This work was supported by the Max Planck Research School on the Life Course.

Eye-tracking data on the processing of privacy protection information on shopping websites suggest that people might decide to ignore information they deem irrelevant across all websites. Alternatively, one could have expected that people start to ignore (seemingly) irrelevant information website by website, still checking privacy information on novel websites while ignoring it on well-known ones (compare Vu et al., 2007). In a similar vein, information on food packages deemed irrelevant seems to be ignored across both well-known and novel food products (e.g., Gaschler, Mata, Störmer, Kühnel, & Bilalic, 2010). A detailed experimental analysis on how information is ignored and how this transfers to novel stimuli therefore seems warranted. Testing whether information reduction occurs stimulus by stimulus or once for all stimuli does not only provide important descriptive information for applied questions. Rather, as detailed below, the one-by-one versus once-for-all question sheds light on two different accounts of strategy change in skill acquisition.

One-by-one versus once-for-all

Lee and Anderson (2001) suggested that much of the learning in tasks involving complex decisions is rooted in a reduction of fixations on task-irrelevant information. Deconstructing learning in air traffic control, the authors ascribed major performance gains to strategy changes that reduce fixations (e.g., skip checking the queue or the weather conditions). This is in line with evidence in the human factors literature that people tend to reduce cognitive effort by discarding irrelevant information from processing (Niessen, Eyferth, & Bierwagen, 1999; Reason, 1990; Underwood, Crundall, & Chapman, 2002). However, there is no agreement on how and why *information reduction* (e.g., Cousineau & Larochelle, 2004; A. Green & Wright, 2003; Haider & Frensch, 1996) takes place. Theories on skill acquisition can be roughly grouped into two broad classes with respect as to how they attempt to account for information reduction. According to one view, the shift to a more efficient strategy is a direct and inevitable consequence of task practice. It is tied to how often the specific material has been practised (e.g., Logan, 1988, 1992; Siegler, 1988). In tasks such as mental arithmetic, participants will show strategy changes one at a time, first with the often-practised task material and later with less often practised material. A voluntary decision is not involved, as applying a less efficient strategy inevitably enforces the shortcut strategy. For instance, according to Logan (1988, 1992), automatic encoding and retrieval of memory traces lead to a transition from a slow calculation-based strategy to faster memory-based responding as a mandatory consequence of task processing. With every trial, the number of memory traces containing past stimuli and responses increases. So does the probability that memory retrieval rather than calculation will determine the response to a problem. The more traces race in parallel for retrieval from memory, the larger the chance that the fastest of these will trigger the response before calculation is completed. We refer to this view as the *bottom-up view*.

The alternative view assumes a contribution of top-down decision components. While automatic (i.e., involuntary and often implicit) learning provides knowledge about regularities in the task environment, it does not mandatorily lead to strategy change. Rather, a voluntary decision is involved. First, participants will generate rule-like knowledge about exploitable regularities in the task material. Then they will *decide* to use this knowledge, reducing cognitive effort by switching to a more efficient task strategy (Haider, Frensch, & Joram, 2005; Rehder & Hoffman, 2005; Sun, Merrill, & Peterson, 2001; Touron & Hertzog, 2004a, 2004b). This decision can change the strategy that is applied to often-practised, less practised, and even novel variants of the task material alike. If implementing the new strategy on the psychomotor level is straightforward (i.e., possible without practice), then the decision can be accompanied by an abrupt change in performance (cf. Haider & Frensch, 2002; Haider et al., 2005). We refer to this view as the *top-down view*. Please note that the views on strategy change are not necessarily mutually exclusive. Prior work has shown that at least some types of strategy change (other than information reduction) can take place without a top-down decision and outside the awareness of participants (cf. Doane, Sohn, & Schreiber, 1999;

Woltz, Gardner, & Bell, 2000). For other conditions a top-down decision is well documented (Touron & Hertzog, 2004a, 2004b). Here we focus on information reduction as one particular variant of strategy change. We acknowledge that in the long run, a theory is needed that accounts for the conditions under which strategy change in skill acquisition takes place according to the bottom-up versus the top-down view (cf. Gray, Sims, Fu, & Schoelles, 2006). As a first step, we develop experimental manipulations and mathematical tools that can help to determine whether a particular type of strategy change is in line with the bottom-up view or the top-down view or shares characteristics of both.

Prominent theories of the bottom-up view (e.g., Cousineau & Larochelle, 2004; Logan, 1988) root strategy change in learning about the specific stimuli in a task (e.g., the specific arithmetic problems practised, rather than arithmetic problems in general). However, some published data patterns suggest equal performance on novel or less practised instances of task material compared to well-practised instances (e.g., Bourne, Raymond, & Healy, 2010; Haider & Frensch, 2002; Harris, Murphy, & Rehder, 2008; Smith, Langston, & Nisbett, 1992; Strayer & Kramer, 1994). Wilkins and Rawson (2010, p. 1134) conceptualized *item-general* practice gains as performance improvements "that accrue to all stimulus tokens of a given type, including both practiced and novel tokens of that type". Therefore, practice gains are item-general to the extent that they transfer to novel tokens. Conversely, they suggested defining *item-specific* gains as improvements that accrue only to the particular tokens that have been practised and not to novel tokens of the same type. Thus, practice gains are item-specific to the extent that gains are greater for practised than for novel tokens.

Studying information reduction: The alphabet verification task

The *alphabet verification task* (Figure 1a; Haider & Frensch, 1996; A. Green & Wright, 2003) has been derived from, and is similar to, alphabet

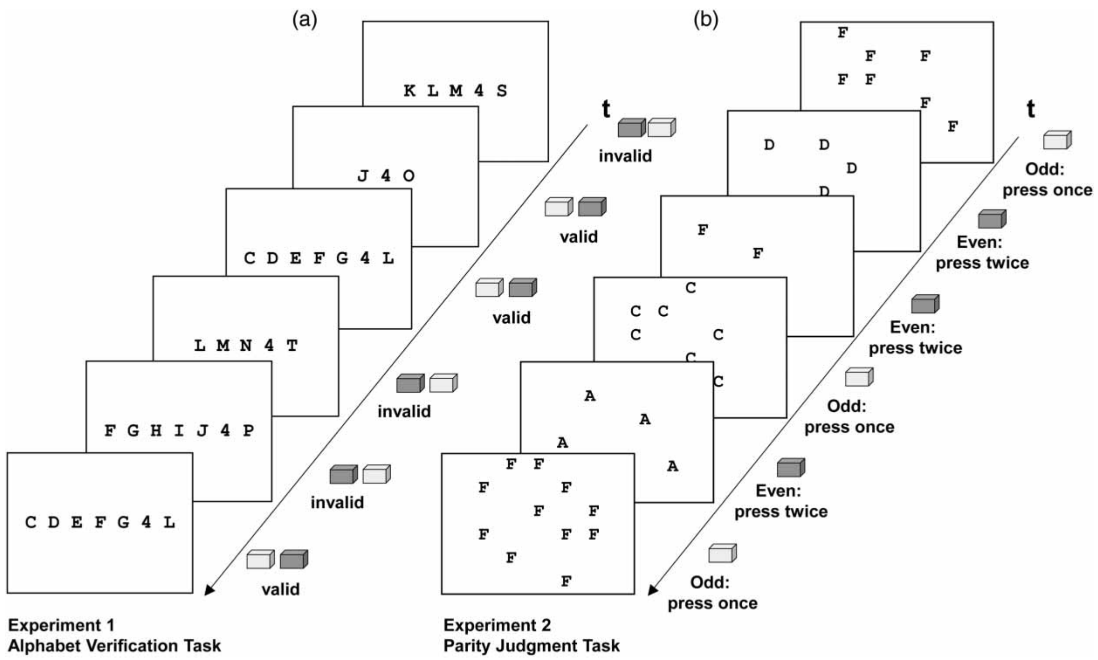

Figure 1. *Task material of Experiment 1 (Panel a) and Experiment 2 (Panel b).*

arithmetic. The latter task has often been employed in work arguing for the data-driven, bottom-up accounts of strategy change in skill acquisition (e.g., Logan, 1988, 1992). In the alphabet verification task, information reduction is possible due to an unannounced regularity in the composition of the alphanumerical strings. A section of the strings can effectively be ignored.

For instance, Haider and Frensch (1999a; Haider et al., 2005) presented participants with letter strings, similar to those of Experiment 1 of the current study. The strings (e.g., C D E F G 4 L) contained a row of letters (e.g., C D E F) ending in a *letter–digit–letter* triplet (e.g., G 4 L). Participants were instructed to verify whether or not the string fully confirmed to alphabetical order. The "4" required participants to skip over four letters. In our example, the string follows the alphabetical order, because C D E F are in alphabetical order, and furthermore there are four letters missing between G and L (H I J K). Violations to the alphabetical order only occurred in the triplet (i.e., the relevant part), while the additional letters always followed the alphabet (i.e., the irrelevant part of the strings). Consistent with the view that information reduction is under voluntary control, the ignoring of the irrelevant part of the strings could be switched on and off by changing speed pressure via instructions (Haider & Frensch, 1999b; Haider et al., 2005). This suggests that participants who knew there were irrelevant aspects in the material did not automatically apply the corresponding processing shortcut. Rather, a volitional decision at least modulated information reduction. Supporting the top-down view further, the new strategy to ignore the irrelevant part of the strings was applied in a transfer block with novel strings (Haider & Frensch, 1996, 2002).

In short, the aforementioned findings allow researchers to conclude that information reduction is item-general—at least if probed *after* reaching proficiency in ignoring the irrelevant aspects of a set of well-known items. However, these findings did not shed much light on the question of how item-general information reduction emerges in the first place—namely, over the course of practice. In an attempt to address this open question, Gaschler and Frensch (2007, 2009) varied the frequency of presentation per block of practice. On the one hand, this manipulation of representation strength led to marked differences in processing times for the *relevant* part of frequent, infrequent, and novel strings. On the other hand, participants ceased to process the *irrelevant* part regardless of the representation strength of the relevant part.

Previous work relied on reaction time (RT) and error-based measures. This made it necessary to aggregate behavioural data at the level of large practice blocks (cf. Gaschler & Frensch, 2007, 2009; A. Green & Wright, 2003). For instance, Gaschler and Frensch (2007, 2009) were able to infer from these aggregated data that information reduction occurred to the same extent for frequent, infrequent, and novel strings. However, they were not capable of specifying the associated *learning curve*—that is, the mathematical function that relates prior exposure to a task to a change in strategy and performance. In our Experiment 1, eye-tracking data were collected to provide a trial-by-trial measure of information reduction that directly disentangled processing of relevant and irrelevant portions of the strings. This measure enabled us to employ competing mathematical models to test whether information reduction would be better accounted for by the one-by-one versus the once-for-all hypothesis (introduced below). In Experiment 2 we explored the basis of generalization between items. We tested whether the application of a new shortcut strategy to well-known and novel items would show similar dynamics to those in Experiment 1, even when there was no spatial regularity in the task material to support generalization.

EXPERIMENT 1: INFORMATION REDUCTION TRIAL BY TRIAL

In Experiment 1, we used eye tracking to investigate whether information reduction is item-general or item-specific as it occurs—or shows characteristics of both variants of strategy change. We combined the frequency manipulation introduced by Gaschler and Frensch (2007) with a string layout that placed the irrelevant part of the letter strings before the relevant one. This required participants to skip over what would normally be read first in left-to-right reading.

Hypotheses and models: One-by-one versus once-for-all

According to the *one-by-one hypothesis*, information reduction depends on the memory strength of the specific items; therefore, strategy change occurs item by item, contingent on the accumulated presentation rate of the items. The performance on the current stimulus is thus best accounted for by a learning curve that depends on the number of times the specific stimulus (alphanumerical string) has been processed in the past.

In contrast, the *once-for-all hypothesis* assumes that information reduction is independent of the memory strength of the particular item encountered. It is the amount of practice in general (irrespective of which specific strings have been practised) that drives information reduction. Therefore, the current performance in a task would be best accounted for by a learning curve that depends on the total amount of prior processing on that task.

Several authors (e.g., Logan, 1988, 1992, for RT; or Lee & Anderson, 2001, for fixation times on irrelevant screen positions) have suggested that in skill acquisition, processing time per trial diminishes as a power function,

$$T = A + BN^{-C}, \qquad (1)$$

whereby A is the asymptote, B is the difference between initial performance and asymptote, C is the rate of learning, and N is the number of trials that have occurred. In our modelling efforts, we will determine how the N-parameter should be specified more precisely: either as the number of prior encounters with the specific item currently presented (one-by-one hypothesis), or as the number of trials that have occurred in general. While the one-by-one hypothesis would be in accordance with the predictions of *instance theory* and related models of the bottom-up view of strategy change (e.g., Logan, 1988, 1992; see above), the once-for-all hypothesis would be compatible with views of skill acquisition incorporating top-down decisions (e.g., Haider et al., 2005; see above).

In order to render the hypotheses as explicit as possible, we will use a reanalysis of published data to illustrate it. We chart competing quantitative predictions for processing times of the irrelevant part of the alphanumerical strings. Figure 2 displays how predicted fixation time on irrelevant parts of frequent, infrequent, and singleton alphanumeric strings should change with practice. In prior work on information reduction, subtraction was used to estimate the time participants spent with the irrelevant part of the strings (cf. Gaschler & Frensch, 2007; A. Green & Wright, 2003). Specifically, RT for items without an irrelevant part was subtracted from RT of items that included an irrelevant part. It is not a given that fixation times are closely linked to RT (Anderson, Bothell, & Douglass, 2004). Therefore, it is time to put the validity of the RT difference measure used in the past studies to the test. To compute predictions, we used reaction time data from the control condition of Experiment 1 in Gaschler and Frensch (2007). In this condition, all participants practised all strings twice per block. In the current Experiment 1, we presented alphanumerical strings three times per block of practice, presented other strings once per block, and, in addition, introduced novel strings in each block. The latter *singletons* were presented in two successive blocks. Thus, to derive predictions, RTs from the past experiment were analysed separately for the *n*th occurrence of each alphanumerical string and were then aggregated according to either the one-by-one hypothesis or the once-for-all hypothesis. Take as an example the prediction of the fixation times for the irrelevant part of infrequent strings in the second block of practice. According to the one-by-one hypothesis, the RT of the second presentation of strings without the irrelevant part would be subtracted from the RT of the second presentation of strings with the irrelevant part. For frequent strings (i.e., three times per block), the prediction for the second block can be generated by averaging the RT of the fourth, fifth, and sixth presentation of the strings and subtracting the

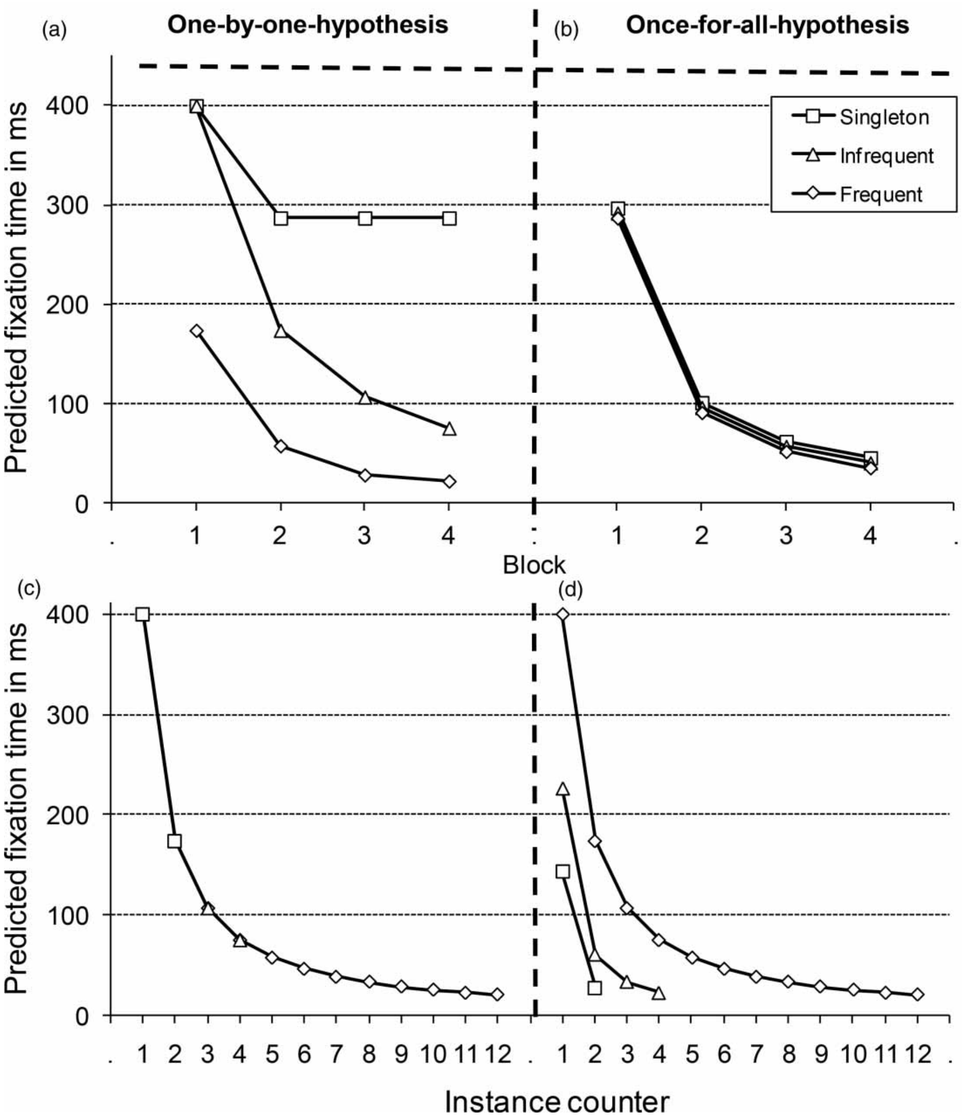

Figure 2. *Predicted fixation time per irrelevant letter of frequent, infrequent, and singleton strings according to the one-by-one hypothesis (left panels) versus once-for-all hypothesis (right panels). In (a) and (b) practice is charted by block (or number of prior encounters with any stimulus) and in (c) and (d) by number of past encounters with the specific stimulus in the current trial. Predictions are based on reaction time data of the control condition of Experiment 1 of Gaschler and Frensch (2007). As each string had been presented eight times during the experiment, power function-based extrapolations were employed to cover the ninth to 12th encounter with the strings.*

processing time of strings without irrelevant part from the processing time of strings with irrelevant part. According to the one-by-one hypothesis, predicted fixation times for the irrelevant part should develop differently across blocks of practice for frequent, infrequent, and singleton strings (Figure 2a). Frequent strings should show faster information reduction than less frequent strings. However, the picture is much simpler according to the once-for-all hypothesis (Figure 2b). The amount of times a specific string has been encountered before should not matter. To summarize, if practice is charted in blocks, then the one-by-one hypothesis predicts an effect of frequency, while the once-for-all hypothesis predicts none. The converse pattern of predictions results if practice is charted by counting the occurrence of the specific alphanumeric strings. In this case, the one-by-one hypothesis predicts no effect of frequency (Figure 2c) while there should be one according to the once-for-all hypothesis (Figure 2d).

Apart from potential differences in processing of frequent, infrequent, and novel stimuli, one further issue has to be taken into account when predicting how fixations change with practice. While Lee and Anderson (2001) proposed that the reduction of fixation time on irrelevant screen positions follows a continuous learning function (the power law), they also provided examples of abrupt changes in fixation times by instructing participants to change their strategy at predefined points in practice. Furthermore, they point out that good power law fits of aggregated data could be based on discontinuous practice gains that occur at different time points for different participants (cf. Haider & Frensch, 2002). According to the once-for-all hypothesis, the major source of efficiency gain would be an item-general abrupt strategy shift that reduces the processing of irrelevant aspects of items to zero. Therefore, a simple step function with two parameters should suffice: the amount of processing prior to the strategy shift, and the item-general trial counter denoting when the shift takes place:

$$T = A_1 \text{if} N < \text{shift},\ T = A_2 \text{if} N \geq \text{shift} \qquad (2)$$

where A_1 is the amount of processing prior to the strategy shift, and N is an item-general trial counter indicating the *change trial*. A_2 denotes the amount of processing after the change trial and can be assumed to be zero if information reduction is fully effective.

Method

Participants

Nineteen university students from Berlin took part in the experiment and were paid €12 (8 female; mean age 25.8 years, $SD = 2.3$). The experiment took place in the laboratories of the Max Planck Institute for Human Development, Center for Adaptive Behavior and Cognition, Berlin, Germany.

Materials and apparatus

The stimuli in the alphabet verification task consisted of 72 alphanumeric strings (e.g., C D E F G 4 L). Half of the 72 alphabetic strings were *valid*, following the order of the alphabet; the other half were *invalid*, deviating from it. Information reduction was possible due to an unannounced regularity in the material. During the four blocks of practice, participants could safely skip over letters outside the triplet, because strings either were valid or contained an error within the letter–digit–letter triplet component. Specifically, a void of five instead of the indicated four letters was present in invalid strings (e.g., L M N 4 T). Strings with violations of the alphabetical order in the letters outside the triplet were only used in the examples employed in the instructions, the 10 training trials prior to the start of the experiment, and the negative transfer block at the end (cf. Woltz et al., 2000). The triplets began with the letters G to R. Each of the valid or invalid letter–digit–letter triplets was displayed together with either no additional letters in front, or an additional two or four letters (e.g., F G H I J 4 P versus H I J 4 P or J 4 P). The crucial experimental manipulation varied string frequency within participants. One third of the strings were shown three times (henceforth: *frequent*), and a second third were presented once per block (henceforth: *infrequent*). The rest of the material was reserved for *singletons*. Each singleton consisted of six strings with the same starting letter in the triplet: a valid and an invalid short, medium, and long string. There were only four singletons. The first singleton was presented in Blocks 1 and 2. The second singleton was introduced in Block 2 and was presented for a second time in Block 3, and so forth. In Blocks 2, 3, and 4, the singletons accounted for 12 of a total of 108 trials (6 of 102 in Blocks 1 and 5—the negative transfer block). The allocation of strings to frequency conditions was balanced across participants with the restriction that all the strings with a given letter in front of the digit were in the same frequency condition.

In the negative transfer block, the part of the strings that had not contained mistakes throughout practice was now no longer safe to ignore. Half of the medium and long frequent, infrequent, and singleton strings that did not follow the alphabetical order were transfer trials. The other half consisted of strings with invalid letter–digit–letter triplets. The number of errors committed on the modified trials (henceforth: *transfer errors*) was used as an additional measure of how little the (formerly) irrelevant part of the strings were processed.

Participants responded by pressing either the "y" or the "comma" key on the second row from the bottom on a standard German PC keyboard. Half of the participants were instructed to use the "y" key to indicate that a string was valid and the "comma" key to indicate that the string was invalid; for the other half, key assignment was reversed.

The strings were presented centrally on the screen in bold Courier New font with a character measuring approximately 0.4° × 0.3°. The character positions were separated by a 5° visual angle in constant spacing with the long strings spanning the whole screen from left to right (30°). A Tobii 1750 eye tracker (17″ TFT screen, 1280 × 1024 pixels; accuracy: 0.5°; drift < 1°) with a sampling frequency of 50 Hz was used. The system works based on corneal reflections. In addition to tracking gaze positions for both eyes, it computes pupil sizes. Participants were seated without head fixation at a distance of 55 cm from the screen. The system was well suited to the comfortable administration of our relatively long and demanding skill acquisition task. While the equipment was tolerant of head movements in a 30 × 16 × 20-cm space, instructions and visual feedback every 30 trials ensured that participants did not move out of the measurement field after they had been calibrated.

Procedure

Participants were instructed to pay attention to the entire string because errors could occur anywhere in the string. Furthermore, they were told to respond as quickly as possible while keeping the rate of errors below 10%. The letter strings used as examples in the instructions and in the 10 practice trials (triplets starting with E and F) were not from the pool of training material.

Each trial started with a fixation cross presented centrally for 1000 ms, followed by a 1000-ms blank interval and then a centrally adjusted letter string. After the manual response was registered, the string was removed, and there was a blank interval of 700 ms before the fixation cross of the next trial appeared. Participants received acoustic feedback when they responded incorrectly (except on the transfer trials in Block 5) and feedback concerning their mean latency and their mean percentage of error upon completion of each practice block. The experiment was completed within approximately 60 min.

Data analyses

Exclusion of data

No RT-based speed–accuracy trade-off was observed in the raw data from Blocks 1 to 4, mean $r(426) = .02$, range $-.15$ to .14, $t(18) = 1.26$, $p = .22$, $\eta^2 = .08$. Three participants with error rates higher than 20% were excluded from all further analyses, resulting in a mean error rate of 7.6% (range: 1.9 to 16.7%). Furthermore, in a preanalysis of eye movement data, 0.14% of the data were excluded because the Tobii system marked the data for both eyes as invalid. Only eye movement data from trials with correct responses were analysed in order to keep close correspondence between the reaction time and fixation data.

Processing of eye movement data

In preparing the fixation data, the seven locations on the screen that could carry a character were defined as *regions of interest* by evenly splitting the distance between the left-most pixels of the adjacent characters. This procedure was validated by a distributional analysis on the horizontal dimension. Next, the regions of interest were grouped into two subsets. On the one hand, we aggregated positions with characters that were relevant (the letter–digit–letter triplet). On the other hand, we aggregated those that could be safely ignored (the prefix letters in strings of length 5 and 7). A fixation was counted whenever five or more consecutive horizontal screen coordinate samples fell within the same region of interest (on the same character location). In addition, the vertical coordinate needed to be within the range of 90% of the distribution of the sample. The latter criterion effectively excluded fixations to the empty top and bottom regions of the screen.

Results

Manipulation checks based on RT, pupil dilation, and fixation times all indicated that the frequency manipulation of the letter strings had successfully influenced the representation strength of the strings. In addition, replicating past work (e.g., Gaschler & Frensch, 2007), aggregated RT and transfer errors suggested that information reduction followed the once-for-all hypothesis (see Appendix A).

Counting practice by item encounter or by trial

The average fixation time on the irrelevant part of frequent, infrequent, and novel strings decreased with block of practice at the same rate and to the same extent (Figure 3a). Using block as an index of practice implies an item-general counter (i.e., the past number of trials with any of the strings of the alphabet verification task). The results resemble the predictions computed based on the once-for-all hypothesis (Figure 2b). An analysis of variance (ANOVA) confirmed a main effect of practice, $F(3, 45) = 33.83$, $MSE = 6393$, $p < .001$, $\eta^2_p = .69$. There was neither a main effect of frequency, $F(2, 30) = 2.11$, $MSE = 1954$, $p = .139$, $\eta^2_p = .12$, nor an interaction of practice and frequency, $F < 1$. Figure 3a indicates that the prior amount of practice with any string (rather than with the specific one presently encountered) determined performance. Apparently, transfer between strings was close to perfect, supporting the once-for-all hypothesis rather than the one-by-one hypothesis or a mixture. For infrequent strings, this suggests that the improvement from one encounter to the next was much stronger than the improvement for subsequent encounters of specific frequent strings. This should be the case according to the once-for-all-hypothesis, because between one encounter and the next with the same infrequent strings there is more occasion for practice with (other) strings than between two subsequent repetitions of a specific frequent string.

In line with this reasoning, additional—and arguably stronger—evidence for the once-for-all hypothesis emerged when the same data were analysed according to how often the specific string currently presented had been encountered before (Figure 3b). Again the pattern resembles the predictions of the once-for-all hypothesis (Figure 2d). The reduction in fixation time from one to the next encounter with an infrequent string was larger than the reduction from one to the next encounter with a frequent string. An ANOVA involving the first four encounters of the frequent strings and all four encounters with the infrequent strings showed a main effect of frequency, $F(1, 15) = 22.69$, $MSE = 5550$, $p = .001$, $\eta^2_p = .6$, and a main effect of encounter, $F(3, 45) = 37.32$, $MSE = 2703$, $p = .001$, $\eta^2_p = .71$. The interaction of frequency and encounter was not robust, $F(3, 45) = 2.56$, $MSE = 3366$, $p = .067$, $\eta^2_p = .15$.

In a follow-up analysis we explored the dynamics of fixations within trials and checked whether the time course of information reduction differed for medium compared to long strings. For instance, one could have assumed that participants start information reduction with the medium strings and only expand it to long strings. They might skip over the (short) irrelevant part and first process the relevant portion of the string, before eventually checking the irrelevant part. Rather, the analyses reported in Appendix A (Figure A3) suggest a uniform time course of information reduction for strings of different length. After some practice even the very first fixations after stimulus onset fell on the relevant part of strings. There was no indication that later fixations were used for checking the relevant part.

Defining practice in the learning curve

We extended the above analysis by learning curve fitting. Lee and Anderson (2001) proposed that the reduction of fixation time on irrelevant screen

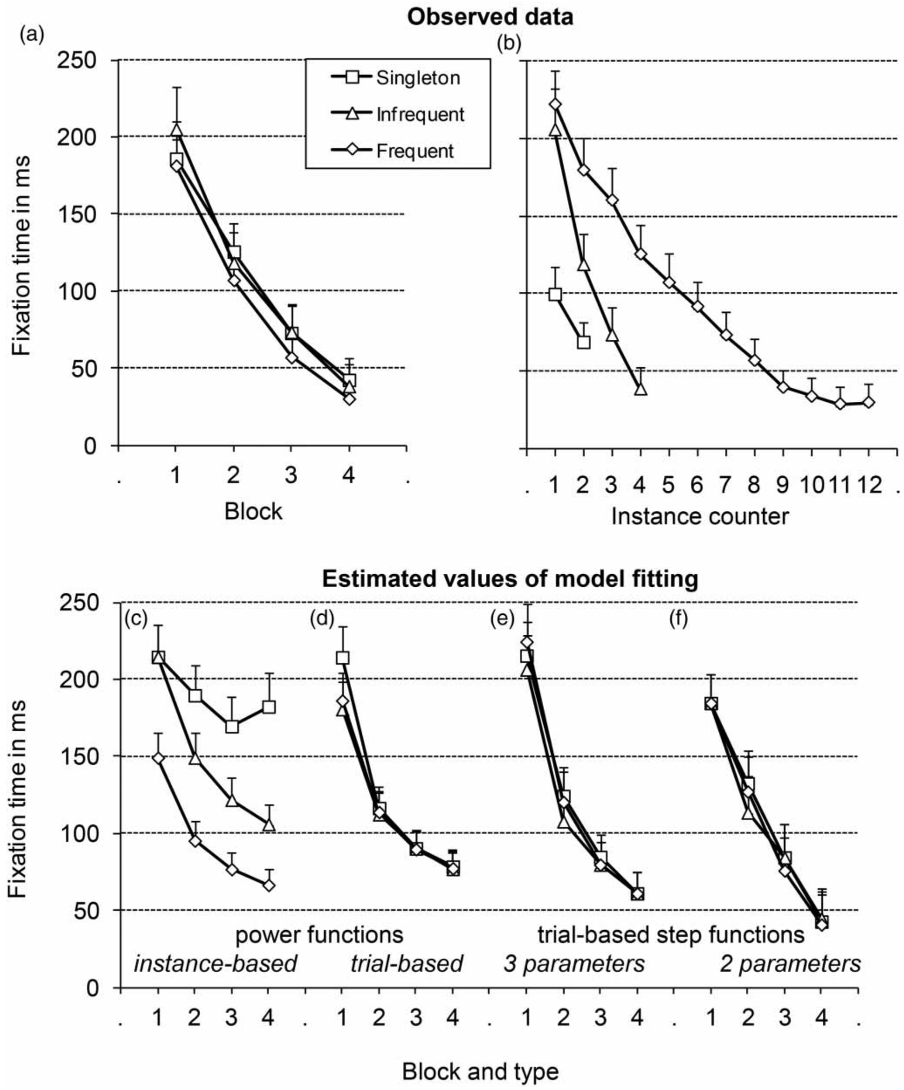

Figure 3. *Mean fixation times per irrelevant character of medium and long strings are charted by block of practice (a) and by number of encounter with each specific alphanumeric string (b). The pattern resembles the predictions of the once-for-all hypothesis shown in Figure 2. Other panels depict the estimated values of (c) the instance-based and (d) the trial-based power function fit, as well as the fit of (e) the three-parameter and (f) the two-parameter trial-based step function. Error bars here and elsewhere display the between-subjects standard error of the mean.*

positions follows the power law. We therefore used fixation time per irrelevant character on medium and long strings as a measure to further specify the learning curve. In a first step, we tested whether the "trial" parameter in the power function (Equation 1) should be defined either as the number of prior encounters with the specific item concurrently present (one-by-one hypothesis) or rather as the number of trials that have passed in general (no matter with which specific items; once-for-all-hypothesis). For each participant the two different power function fits were calculated and compared.[1] The fitted estimates for frequent, infrequent, and singleton trials overlapped completely for the (item-general) trial counter (Figure 3d—just as the observed values in Figure 3a do). However, fitted estimates markedly diverged for the instance counter (Figure 3c). That is, the instance-based model incorrectly ascribed singleton and infrequent strings too much fixation time on the irrelevant part of the strings.

For only two of the 16 participants, the root mean square error (*RMSE*) was smaller when the instance counter was used as opposed to the trial counter, $M = 113.43$ versus 107.26 ms deviation, $t(15) = 3.64$, $p = .002$, $\eta^2 = .47$. This is consistent with the once-for-all hypothesis of information reduction. The size of the effect is quite surprising given the substantial correlation of trial counter and instance counter ($r = .72$, averaged across participants, range .69 to .74). When analysed separately for the frequency classes, the disadvantage of the fit based on the item-specific instance counter as compared to the item-general trial-based fit was particularly pronounced in the case of the singleton items ($M = 139.36$ versus 106.49 ms deviation), $t(15) = 6.15$, $p = .001$, $\eta^2 = .72$.

Continuous or abrupt learning curves for information reduction?

The above analysis supported the once-for-all hypothesis, which is consistent with a contribution of a top-down decision to strategy change. As such strategy change has been linked to abrupt performance gains (e.g., Haider & Frensch, 2002; Opfer & Siegler, 2007), we conducted three model tests to check whether information reduction on the trial-by-trial level would be better captured by a step function or a continuous learning curve (such as the power function reported by Lee & Anderson, 2001).

Model tests I: Fitting fixation data. To assess the evidence for continuous versus discontinuous reduction of fixation time on irrelevant letters, we fitted step functions and compared them to the (item-general) trial-based power law fits. If participants abruptly switch from checking the irrelevant letters to no longer fixating them, then the data should be captured well by a step function with a high fixation time prior to the strategy change and a low fixation time after the strategy change.

All possible step functions were generated for each participant, and the one with the best fit was chosen. From the first to the last, each trial was considered as the potential point in practice where the participant may have switched from fixating on the irrelevant letters to ignoring them. The compound *RMSE* for two zero slope lines, one before this trial and one beginning with this trial, was compared. We used all trials with irrelevant characters—that is, long and medium strings. The trial that led to the smallest overall *RMSE* was taken as the point in practice where the strategy change occurred (see Rickard, 2004, for a similar approach). The three-parameter step function (higher intercept, lower intercept, change trial; Equation 2) led to smaller *RMSE*s in 11 out of 16 participants than the trial-based three-parameter power function, Equation 1; deviation, $M = 101.5$ ms, $t(15) = 2.11$, $p = .052$, $\eta^2 = .23$. Thus, with an equal number of free parameters, the nonaggregated data tended to be better fitted by a step function than by a power function. An analysis of the fixation times prior to and after the change trial suggested that the reduction of the course of practice to two straight asymptotes and a shift point was warranted. Specifically, we computed independent power function fits for the fixation times prior to and after the change trial for each person. The *RMSE*s of the straight asymptotes (one parameter) were just 4 and 3 ms higher than those of the respective three-parameter power function fits. At the same time, if aggregated across participants, the estimated values from the step function model mirrored the aggregated practice data well (Figure 3e, f; cf. Haider & Frensch, 2002, for a discussion of aggregation artefacts in skill acquisition data).

[1]Curve fitting for the step functions was performed with a custom-written algorithm in Pascal based on least square estimates. For fitting power functions, the curvature parameter was restricted to negative, and the asymptote was restricted to ≥0 ms. We used the sequential quadratic programming algorithm included in SPSS.

Note that with fewer free parameters the step function was equally successful as the power function. The fits of a two-parameter step function resulted in an *RMSE* equal to the fit of the three-parameter power function ($M = 107.99$ ms). Given that many participants reduced fixations on most of the irrelevant letters to zero, just two free parameters (one for the higher intercept, one for the change trial) seemed sufficient.

Model tests II: Predicting fixation data. Further credence is lent to a model by accurate prediction (i.e., without any further parameter adjustments; cf. Marewski & Olsson, 2009; Pitt, Myung, & Zhang, 2002; Roberts & Pashler, 2000; Wagenmakers, 2003). Above we reported that the fit of the step function was (a) in tendency better than the one of the power function with an equal number of free parameters and was (b) identical when a step function with just two free parameters competed with a three-parameter power function. Importantly, the advantage of the step function over the power function was not limited to fitting but also extended to predicting new data in a cross-validation: First we derived curve fits for both of the two functions separately for each of the four frequent items (48 data points each per participant). Next we checked how well the parameters derived for one frequent item could predict the performance on each of the other three items. The step function yielded better prediction results than the power function ($M =$ 107.59 ms versus 117.47 ms; averaging over all combinations of predictions), $t(15) = 2.84$, $p = .012$, $\eta^2 = .35$. Note that model mimicry could not account for the advantage of the step function (i.e., the phenomenon that some functions are flexible enough to capture the output of other functions; see Appendix B).

Model tests III: Individual analyses and change trials. Figure 4a illustrates why the step function was in some cases better in fitting and predicting the data than the power function. The figure displays

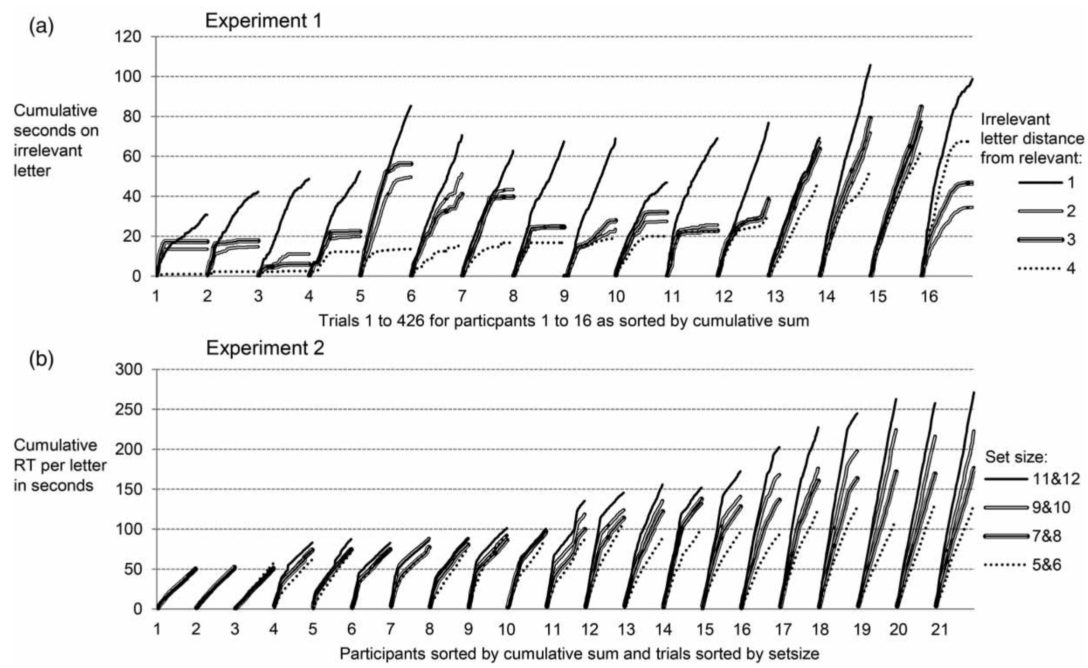

Figure 4. *Cumulative fixation times for the time spent on each of the four irrelevant characters on long strings in Experiment 1 (Panel a). Participants are ordered by the overall time spent. Panel b shows the cumulative time curves for Experiment 2 separately for each of the four large set sizes. RT = reaction time.*

the cumulative fixation times over all trials for each of the four irrelevant characters of the long strings for each of the participants. Incompatible with a continuous learning function, some participants show a shift in slope. Equally incompatible with the power function (cf. Haider & Frensch, 2002), the largest reduction in time spent on the irrelevant characters often seemed to take place late in practice (rather than early in practice). Therefore, even for participants who did not show a marked shift in the slope of the cumulative curve, the power function did not fit well. Differences in fixation between irrelevant letters close to versus remote from the relevant string part seemed in part to explain why overall the reduction in fixations on irrelevant letters was not abrupt for some participants. Reduction was stronger and more abrupt for the irrelevant letters located further apart from the relevant part of the strings. Therefore, averaging across the different letters might have masked abrupt changes in some cases.

The above analyses suggest that participants apply information reduction in a once-for-all manner, and some show an abrupt strategy change. Therefore, the change trials in step functions independently derived for singleton, infrequent, and frequent strings for each person should fall close together. Averaged over participants, the change trials for frequent ($M = 242$), infrequent ($M = 247$), and singleton ($M = 263$) trials did not differ from each other ($F < 1$). Figure 5a depicts the change trials for each person. It is evident that there is large between-person variability. For some participants, change trials occurred early, while others changed late in practice. Importantly, within persons the independently derived change trials fell close together in many cases. Cronbach's alpha for the three change trials was .82. The correlation between change trials derived for frequent and the one derived for infrequent strings was $r(16) = .76$ (.57 and .53 for the other pairings, all $p\text{s} < .035$). High intraindividual consistency paired with high interindividual variability in change trials was also observed on the item level (compare Touron,

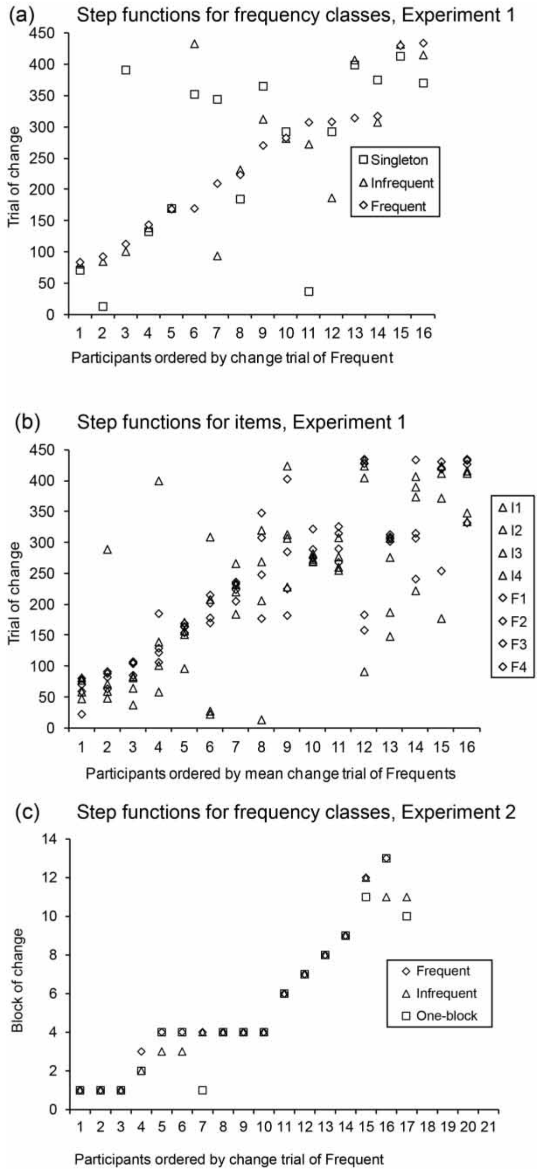

Figure 5. *Panel a shows the locations of change trials (y-axis) for frequent, infrequent, and singleton strings per participant (x-axis) of Experiment 1. Panel b allows comparison of the independently derived step functions for the four frequent (F1 to F4) and infrequent items (I1 to I4) of each person. Panel c shows the change block analysis for Experiment 2. It depicts the block in which reaction time per letter was reduced below 50 ms—calculated independently for each participant and the frequent, infrequent, and one-block-only items.*

2006). For this we fitted two-parameter step functions independently to the four frequent and four infrequent strings of each person. The average correlations among the eight change trials was $r(16) = .77$, $p < .001$, Cronbach's alpha = .96.

Discussion

Analyses of Experiment 1 suggested that information reduction occurs in a once-for-all manner for different alphanumeric strings. With the specific experimental setup employed, we did not obtain evidence that one-by-one strategy change might be at least partially involved as well. Rather, participants reduced fixation on irrelevant parts of frequent and infrequent stimuli at the same rate and to the same extent with practice. Once acquired, information reduction was also applied to novel items. Modelling the learning curve, the improvement in performance over trials of practice was therefore not captured well if "trial" denoted the number of prior encounters with the specific item currently present. Such an interpretation of "trial" would have been suggested by the one-by-one hypothesis of information reduction and theories in accordance with the bottom-up view of strategy change (e.g., Cousineau & Larochelle, 2004; Logan, 1988, 1992). Performance improvement based on reduction of fixation time on the irrelevant part of the alphanumeric strings was better captured by a learning curve that specified "trial" as the number of *any* strings presented so far. This is consistent with the once-for-all hypothesis of information reduction and the view that top-down decisions are involved in strategy change (e.g., Haider & Frensch, 2005; Rehder & Hoffman, 2005; Sun et al., 2001; Touron & Hertzog, 2004a, 2004b).

The validity of our support for the once-for-all hypothesis of information reduction was demonstrated by manipulation checks indicating that we had successfully varied representation strength: Participants processed the *relevant* part of frequently versus infrequently presented strings differently. In line with top-down theories of strategy change, we furthermore observed abrupt strategy change for at least some of the participants. For such shifts to occur, the psychomotor routines implementing the cognitive strategy must be changed quickly and fully when the strategy changes. This might provide one explanation for why in many cases fixations on the irrelevant letter bordering the relevant part of the strings were reduced less or not at all. Furthermore, some participants might have continued to use the irrelevant letters close to the letter–digit–letter triplet as cues to start retrieval of subsequences of the alphabet in order to process the relevant part of the strings. Neither for participants with nor for those without abrupt change was there a tendency that information reduction might occur earlier for frequent than for infrequent alphanumeric strings.

In Experiment 1, we employed an extension of the alphabet arithmetic task. The original version of this task had been used in studies supporting one-by-one strategy change with a learning curve following the power law (e.g., Logan, 1988, 1992). According to the instance theory, strategy change should occur earlier for frequent and later for infrequent task material. Performance should improve gradually and show the largest gains at the beginning of practice. Experiment 1 did not provide supportive evidence for the bottom-up view. Once-for-all strategy change was evident in that the irrelevant part was ignored to the same extent in frequent and infrequent alphanumeric strings. In addition, information reduction took place abruptly for at least some of the participants. Furthermore, the largest performance gains often occurred rather late in practice. This limits the extent to which skill acquisition theories employing the power law of practice can account for the data, because they predict that the largest gains in performance should occur early in practice (e.g., Logan, 1988, 1992) involving the reduction of fixations on irrelevant aspects of tasks (Lee & Anderson, 2001).

In the current experiment, we combined a frequency manipulation with trial-by-trial assessment of information reduction. Prior work on information reduction that argued for the top-down view of strategy change used aggregated data (Gaschler & Frensch, 2007, 2009; see also A. Green & Wright, 2003) and relied on block-based estimates rather than on a learning curve charted on a single trial level. Apart from lack of temporal precision, such data were ambiguous. For instance, the longer times that might be required for processing the relevant part of infrequent and novel strings could have been traded off against the times required to process the irrelevant part. Therefore the results of Experiment 1 provide much stronger support for the once-for-all hypothesis than aggregated reaction times in previous research.

EXPERIMENT 2: ONCE-FOR-ALL WITH RANDOM SCREEN POSITIONS

On the one hand, information reduction occurred once for all stimuli under the conditions tested in Experiment 1. On the other hand, this observation does not necessarily lead to the conclusion that information reduction is based on top-down strategy change. Bottom-up attention learning might also lead to once-for-all strategy change. If participants learned what (relative) position of the strings to attend to and/or fixate (e.g., around the digit), this could allow for generalization of information reduction as well—a top-down decision would not necessarily have to be involved. It is conceivable that learning of the relevance of screen positions leads to information reduction generalizing across frequent, infrequent, and novel items—because they are all displayed in the same spatial layout. Such a bottom-up attention learning perspective could, for instance, be derived from the attention learning model in Kruschke (2001) or the work of Logan (2002) combining attention and learning (see also C. S. Green, Pouget, & Bavelier, 2010).

Work with the alphabet verification task has taken first steps to explore whether screen position learning is a major contributor to information reduction. Haider and Frensch (1999a) demonstrated that information reduction does take place when strings starting versus ending with the irrelevant part are mixed on a trial-by-trial basis. In Experiment 1, strings were presented in a centred manner, so that the absolute screen position of the relevant and irrelevant string parts were different for long than for medium strings. These results suggest that people learn to ignore irrelevant information in visual tasks that provide at least some variability in where on the screen the irrelevant information is located. As a next step, we tested whether once-for-all strategy change would also be found in a task with random placement of items, which prevents the learning of to-be-attended positions.

For this we developed a parity judgment task (Figure 1b), which allowed exploring whether findings from the alphabet verification task would be mirrored by similar patterns of strategy change in a different task format. On the one hand, in our parity judgment task processing could be simplified by applying a shortcut strategy. On the other hand, no opportunity was provided to learn where spatial attention and fixations should be directed on the screen to allow for the shortcut. Participants in Experiment 2 were instructed to determine in each trial whether the number of instances of a letter randomly scattered over the screen was odd or even. The task required participants to identify the identity of the letter, as they had to press the according key. However, each trial featured only instances of one specific letter. It was therefore not necessary to fixate each of the instances in order to solve the task. Therefore, RT as a traditional measure in enumeration experiments (e.g., Watson, Maylor, & Bruce, 2005) seemed more appropriate than fixation counts. Reaction time should monotonically increase with set size—at least early in practice, when participants would still need counting to determine the response for large stimulus sets. However, unbeknownst to the participants, the correct answer was fixed for trials with large stimulus sets. Once participants acquire the regularity in the task material and apply it to simplify task processing, short response times should become possible even for large stimulus sets.

Method

Participants

Twenty-one university students from Berlin took part in the experiment and were paid €10 (18 female; mean age 25.3 years, $SD = 5.7$). The experiment took place in the laboratories of the

Psychology Department of Humboldt-Universität Berlin.

Materials and apparatus
On each trial, participants were shown a display with one to 12 instances of a capital letter—for example, 9 times the letter "E" in white against a black background, 0.7° × 0.5° each, on a 17-inch CRT screen operating at a resolution of 800 × 600 pixels. In order to avoid cluttering, the instances of the letter were randomly placed based on a 6 × 6 matrix in an array of 6.8° × 6.8° at the centre of the screen. Participants were instructed to determine whether the number of instances of the letter on the screen was odd or even. Counterbalanced across participants, the key corresponding to the letter used in the current trial was to be pressed either once or twice to indicate whether the number of letters presented was odd or even. For instance, in a trial with multiple instances of the letter E on the screen, the participants were to press the "E" key once or twice depending on whether the number of "E"s was odd or even. After a key was pressed, there was a 300-ms deadline to register a second press of the same key. In valid trials with double key presses we used the RT of the first key press for later analysis (cf. Wenke, Gaschler, & Nattkemper, 2007). If the response was incorrect, a 200-Hz tone was played as error feedback during the first 250 ms of the 500-ms response–stimulus interval.

We used the letters in the upper two rows of a German keyboard (Q, W, E, R, T, Z, U, I, O, P, A, S, D, F, G, H, K). For each participant, one of these letters was randomly selected to be the frequent one, being used on 36 of the 60 trials in each of 14 practice blocks. A second letter was present in 12 of the trials of each block (infrequent letter). For the remaining 12 trials per block, a different letter was randomly chosen without replacement for each block from the remaining letters (one-block letter).

Apart from the negative transfer block, there was a regularity that made it possible to quickly give correct answers without counting the number of instances of a letter in a given trial. For half of the participants, the material was arranged in a way that the number of instances of the letter was even whenever it was larger than four. For the remaining participants, a cloud of more than four letters was always of an odd number. Each participant was presented with an equal number of stimuli commanding an "odd" versus an "even" response. Under the constraint of the task regularity described above, each number of instances of a letter was presented with equal frequency in each practice block. In the negative transfer block, all set sizes were presented with equal frequency. One third of the trials were transfer trials. They tested whether participants would perform according to the (former) task regularity. For instance, Participant 1 received a transfer trial with eight instances of the letter "W", while in Blocks 1 to 14 the participant would have received an odd number or a number smaller than 5.

Procedure
Participants started the experiment with written instructions asking them to determine whether the number of instances of the letter on the screen was odd or even. No reference to potential regularities or shortcuts was made. The experimenter than watched the first five trials to make sure that the participants had properly understood the instructions and were able to execute the double key presses. After 14 practice blocks, participants received the negative transfer block (without special instruction). Auditory error feedback was disabled for the transfer trials. Finally the experimenter conducted a short structured interview inquiring about explicit knowledge concerning the regularity in the tasks.

Results

Screening of the data suggested that there was no speed–accuracy trade-off. Error trials tended to be slower than correct trials, $r(840) = .07$, and the error rate was low ($M = 5.8\%$). The manipulation check indicated that the frequency manipulation was effective (see Appendix A, also for transfer errors). In the postexperimental interview, participants showed exhaustive knowledge of the regularity in the task material that could be exploited.

All but one correctly reported which categorization ("odd" or "even") was generally valid when more than three instances of a letter were presented. Furthermore, all participants could report correctly which letter was presented most frequently and which one was second most frequent.

RT per letter—From counting to shortcut strategy

Screening of the data indicated that there was a linear increase in reaction time with the number of letters on the display. We therefore used linear regression to determine RT per letter (independently for trials with frequent, infrequent, and one-block-only letters). RT per letter should indicate whether participants were processing based on counting versus based on the shortcut strategy. As depicted in Figure 6a, in line with the once-for-all hypothesis, RT per letter was very similar for all frequency variations and decreased at the same rate per block of practice. This impression was supported by a 3 (frequency of presentation: frequent, infrequent, one-block-only) × 14 (block: 1–14) repeated measures ANOVA. There was a main effect of block $F(13, 260) = 15.78$, $MSE = 22{,}332$, $p = .001$, $\eta^2_p = .44$, while there was neither a main effect of frequency ($F = 1.83$), nor an interaction of frequency and block ($F = 1.03$).

In the block-based analysis, the evidence for item-general strategy change is based on a null effect. However, frequency of presentation has a pronounced effect on RT per letter, if the same data are charted with the item-specific trial counter on the x-axis (i.e., how often the letter used for the current trial had been used before; Figure 6b). As in Experiment 1, infrequent stimuli showed larger performance gains from one encounter to the next than frequent stimuli. An ANOVA involving the first 14 encounters of the frequent stimuli and all 14 encounters with the infrequent stimuli showed a main effect of frequency, $F(1, 20) = 19.41$, $MSE = 26{,}639$, $p = .001$, $\eta^2_p = .49$, a main effect of encounter, $F(13, 260) = 11.24$, $MSE = 7795$, $p = .001$, $\eta^2_p = .37$, and an interaction of frequency and encounter, $F(13, 260) = 4.32$, $MSE = 10{,}900$, $p = .001$, $\eta^2_p = .18$. The finding that RT per letter seemed to increase within the first block of practice was unexpected. Probably, other sources of variability mask the impact of set size on RT. Presumably participants were still becoming accustomed to the instructed task during the first block, which included double key presses for half of the answers.

Averaged over participants, the data suggested a gradual decrease in processing time with practice. However, a look at the learning curves of individual participants again showed that changes in performance were rather abrupt in at least some of the cases (compare Figure 4b). For three participants the

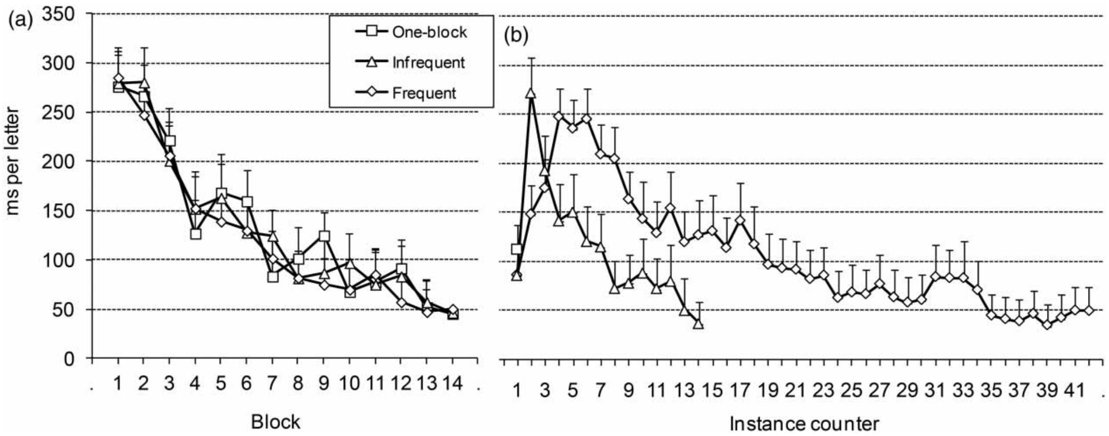

Figure 6. *Both panels show the average reaction time per letter for frequent, infrequent, and one-block-only items of Experiment 2. The data are displayed for either (a) the 14 practice blocks or (b) the number of past encounters with the specific stimulus.*

cumulated RT did not differ at all for larger versus smaller stimulus sets, suggesting that these participants had already discovered and applied the shortcut strategy in the first block. For many other participants the shortcut strategy seemed to set in at the same point in practice at least for the largest set sizes. In the respective cumulative curves, slopes became parallel after a phase in which larger set sizes led to slower reactions and hence to steeper slopes in the cumulative graph.

Four participants did not seem to apply the shortcut at all. For them, RT per letter did not diminish below 50 ms over the course of practice. The remaining participants showed a pronounced reduction in RT per letter from one block to the next when they crossed the 50 ms criterion ($M =$ 219 ms; three were below 50 ms already in Block 1). This reduction was consistent across the levels of the frequency manipulation. As depicted in Figure 5c, the first block with an RT per letter of less than 50 ms was identical for frequent, infrequent, and one-block-only letters for most participants. Correlations for the change points across participants ranged between $r(21) = .97$ and $r(21) = .98$, $p\text{s} < .001$; Cronbach's alpha = .99.

Discussion

In Experiment 2 we observed once-for-all strategy change in a task, which did not allow for learning of screen positions. Please note that one-by-one strategy change was granted a fair chance in this newly developed task. Participants had to give an item-specific response for each item anyway. They pressed the key of the specific stimulus letter. Thus, the frequency manipulation involved stimuli and responses. Therefore, it should have been feasible, in principle, to change to the shortcut strategy early for frequent and later or never for infrequent or novel stimuli. However, strategy change occurred for frequent and infrequent items at the same rate in practice. Once participants had acquired and applied the shortcut option, they also transferred it to novel stimuli.

Initially participants had to use counting in order to determine whether the number of instances of a letter in a given trial was odd or even. With practice on the task material, many stopped following the instructions and instead exploited the unannounced regularity in the task material. There was a fixed correct response for any display with more than four letters, and it was therefore not necessary to fully process the letter display (e.g., by counting). Most participants eventually discovered and exploited this regularity in a once-for-all manner. As screen positions were random, and responses were stimulus-specific, the data make it harder to argue that attention-learning accounts (e.g., Kruschke, 2001; Kruschke, Kappenman, & Hetrick, 2005) can explain the once-for-all simplification of task processing observed in the current experimental setups. In these theories, attention may be allocated to or drawn away from specific features or positions. Neither of these options would have helped to produce the item-general strategy change observed. Rather, participants probably used the *amount* of processing as the basis for learning about the task regularity (i.e., one response is always required in the case of high counts).

GENERAL DISCUSSION

Our two experiments had two goals. First, we tested whether information reduction occurs in a once-for-all manner or in a one-by-one manner using eye movements. Second, we explored how and why this might be the case. Prior work on other types of strategy change in skill acquisition has shown that qualitative changes in processing strategies can in some cases take place without a top-down decision and outside the awareness of participants (cf. Doane et al., 1999; Woltz et al., 2000). However, other work has underlined the relevance of top-down decisions, which imply once-for-all strategy change in other variants of skill acquisition tasks (Touron & Hertzog, 2004a, 2004b). We suggest that the long-term goal should be a theory that accounts for the specific conditions under which strategy change in skill acquisition conforms to the one-by-one hypothesis, conforms to the once-for-all hypothesis, or shows characteristics of a mixture. Thus, for information reduction in skill acquisition, the relevant question might become *when* rather than *whether* it incorporates a top-down decision or exclusively relies on bottom-up learning mechanisms alone (cf. Gray et al., 2006; Marewski & Link, 2014; Marewski & Schooler, 2011, for work on strategy selection that might offer a starting point). With the current work we provide a first step towards such a theory. Frequency variation and covert trial-by-trial strategy assessment employed provide the means to detect both the impact of top-down decisions and the consequences of bottom-up learning processes. Thus, such techniques might be used to map under which conditions strategy change in skill acquisition occurs once for all, occurs one by one, or shows a mixture.

Though participants were instructed to exhaustively check alphanumeric strings as to whether or not they followed the alphabetical order, with practice many solved the task other than instructed. They stopped to check the part of the strings that effectively did not contain mistakes. Under the conditions tested in this study, information reduction occurred in line with the once-for-all hypothesis. We found no evidence for an (additional) influence of the number of prior encounters with the specific stimulus. On the one hand, this might be surprising given other well-documented cases of bottom-up strategy change (other than information reduction—see above). On the other hand, our findings are in line with observations from applied settings involving information reduction (e.g., Gaschler et al. 2010; Vu et al. 2007). In Experiment 1, we studied how fixations on irrelevant aspects of alphanumeric strings were reduced with practice. With detailed trial-wise analyses and comparison of mathematical models we documented that information reduction occurs in line with the once-for-all hypothesis in the alphabet verification task. With practice, processing of irrelevant information was reduced at the same rate for frequent and infrequent stimuli. Once simplification of task processing took place, it was also transferred to novel stimuli. Thus, once acquired with some stimuli, it was applied to all. In addition, for at least some participants the strategy change seemed to take place abruptly. Despite substantial differences in the processing of the relevant part of frequent, infrequent, and singleton strings, the irrelevant part was ignored independently of how often the current string had been presented before. In line with the once-for-all hypothesis, information reduction on novel or irrelevant strings did not lag behind the reduction of fixations on the irrelevant letters of frequent strings.

Theories advocating the bottom-up view of strategy change in skill acquisition (e.g., Logan, 1988, 1992) and information reduction in particular (e.g., Cousineau & Larochelle, 2004) would predict that strategy change should depend on the frequency with which the specific strings have been encountered (one-by-one hypothesis). It should take place earlier for frequent material, and late or never for infrequently encountered strings. This is because these theories route performance gains in a transition from (a) providing the answer to a problem by algorithmic processing to (b) retrieving the answer to the specific problem from memory. Please note that this implies that these theories are consistent with the fact that people can use rules and can therefore process task material in an item-general way—mainly at the beginning of practice. However, we argue that this way of incorporating rule-based processing does not provide an account for our data. For instance, Logan (1988) argues that with increasing practice, the chance increases that on a given trial, memory retrieval of the correct answer to an alphabet arithmetic problem is faster than the algorithmic solution. It is thus predicted that task processing is rule based and item-general in the beginning of practice, when there are not yet any memory traces that could win the retrieval race against the item-general counting algorithm. Later in practice, task processing should become more and more item-specific. It is noteworthy that the task in Experiment 1 was very similar to the alphabet arithmetic task used by Logan, and practice-related changes in the processing of the *relevant* part of the alphanumerical strings were indeed consistent with the instance theory and related approaches. However, information reduction (skipping processing of the irrelevant part) occurred with practice without a change from item-general to item-specific processing.

Experiment 2 dealt with explanations for once-for-all strategy change, which the bottom-up view could put forward. In particular, a generalization across different alphanumeric strings should be feasible by learning at which (relative) screen positions attention and eye fixations should be directed. For instance, Lee and Anderson (2001) suggested that much of the learning in tasks involving complex decisions is rooted in a reduction of fixations on task-irrelevant information. However, Experiment 2 documented that once-for-all simplification of task processing can occur even if there are no spatial regularities that learning of attention or eye movements could possibly exploit. Thus once-for-all simplification of task processing does not seem to depend crucially upon a spatial learning process that inherently supports generalizations across specific variants of material presented at consistent screen positions. Rather, information reduction seems to be item-general, because participants decide that they no longer want to check the characters they deem irrelevant in the alphabet verification task and as a consequence no longer focus on those characters (cf. Rehder & Hoffman, 2005).

We observed close to perfect transfer of information reduction between well-known and less well-known alphanumeric strings in the alphabet verification task. Also, our participants could report on the regularities and shortcuts in the tasks at the end of the experiments. Both points distinguish our findings from prior research on skill acquisition in which (a) transfer across different items within a task was only partial, and (b) participants were unaware of strategy application and could not control it (cf. Doane et al., 1999; Woltz et al., 2000). Other authors studied implicit learning of spatial layout in tasks that prevented participants from becoming aware of regularities (e.g., Jiang & Song, 2005). They documented surprising cases of item-specific learning. If even spatial learning can be item-specific, it seems even more surprising that information reduction on alphanumeric strings was item-general. Currently, we can only speculate that the availability of general verbal labels as well as awareness of the task regularity might be crucial for strategy change in line with the once-for-all hypothesis. However, it is important to point out that in our tasks item-specific strategy change was possible in principle. We observed once-for-all strategy change even though several aspects of the design even fostered one-by-one strategy change. In Experiment 2, we even required participants to press a specific key for each item, one of them very frequently and others infrequently. Yet, participants applied the shortcut at the same time for all items. Participants in the alphabet verification task (Experiment 1) could in principle have applied information reduction to frequent but not to infrequent stimuli. Participants were varying their processing strategy for the relevant part of the strings from trial to trial, depending on string presentation frequency. As a quantitative distinction, the relevant part of infrequent strings was processed much slower than the relevant part of frequent strings. As a qualitative distinction, the relevant part of frequent strings was most likely often processed by memory retrieval, whereas the relevant part of infrequent strings was processed by counting through the alphabetical order. Memory retrieval of solutions to alphabet arithmetic problems could be applied only to problems that a participant had encountered at least once before (cf. Touron, 2006). Differences in processing of the relevant part of frequent versus infrequent items could have provided the trigger for selecting information reduction on frequent but not on infrequent items. Our manipulation checks on the impact of the frequency variation on processing of the relevant part of stimuli showed strong frequency effects. Participants were thus able to parse alphanumeric strings and process the relevant part of stimuli while ignoring the irrelevant part. Follow-up experimentation furthermore suggests that processing times for the relevant part of the frequent stimuli are already close to the asymptote by the end of training in the current experiments.

In summary, we contrasted predictions of the bottom-up view of strategy change in skill acquisition (e.g., Cousineau & Larochelle, 2004;

Logan, 1988, 1992) with those of the top-down view (e.g., Haider & Frensch, 2005; Rehder & Hoffman, 2005; Sun et al., 2001; Touron & Hertzog, 2004a, 2004b). The former view suggests that strategy change is a learning phenomenon. The latter suggests that it is a learning-plus-decision phenomenon. According to the top-down view, strategy change is not a direct inevitable consequence of task practice. Rather, people decide to apply/not to apply the incidentally acquired knowledge about a regularity in the task material for a shortcut. This decision can be a general one, including well-known as well as novel stimuli alike and can lead to abrupt changes in performance. Even though skill acquisition tasks usually do not explicitly prompt decisions, participants incidentally learn about regularities in the task material and through this discover that there is a decision to be made: to apply this knowledge for the simplification of task processing or, alternatively, to go on processing the task as originally instructed by the experimenter. Taken together, the present data suggest that information reduction can be based upon a voluntary decision to exploit incidentally acquired rule knowledge about task regularity for simplification of task processing. Therefore, future research can extend volition intervention approaches (e.g., Bayer, Gollwitzer, & Achtziger, 2010; Fujita & Roberts, 2010; Job, Dweck, & Walton, 2010) to strategy changes aiming at the simplification of task processing.

Original manuscript received 23 January 2013
Accepted revision received 23 June 2014
First published online 6 October 2014

REFERENCES

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval: Limits of the eye-mind hypothesis. *Psychological Science*, *15*, 225–231.

Bayer, U. C., Gollwitzer, P. M., & Achtziger, A. (2010). Staying on track: Planned goal striving is protected from disruptive internal states. *Journal of Experimental Social Psychology*, *46*, 505–514.

Bourne, L. E., Jr., Raymond, W. D., & Healy, A. F. (2010). Strategy selection and use during classification skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 500–514.

Cousineau, D., & Larochelle, S. (2004). Visual-memory search: An integrative perspective. *Psychological Research*, *69*, 77–105.

Doane, S., Sohn, Y. W., & Schreiber, B. (1999). The role of processing strategies in the acquisition and transfer of a cognitive skill. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1390–1410.

Fujita, K., & Roberts, J. C. (2010). Promoting prospective self-control through abstraction. *Journal of Experimental Social Psychology*, *46*, 1049–1054.

Gaschler, R., & Frensch, P. A. (2007). Is information reduction an item-specific or an item-general process? *International Journal of Psychology*, *42*, 218–228.

Gaschler, R., & Frensch, P. A. (2009). When vaccinating against information reduction works and when it does not work. *Psychological Studies*, *54*, 42–53.

Gaschler, R., Mata, J., Störmer, V., Kühnel, A., & Bilalic, M. (2010). Change detection for new food labels. *Food Quality and Preference*, *21*, 140–147.

Gray, W. D., Sims, C. R., Fu, W., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*, 461–482.

Green, C. S., Pouget, A. & Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games. *Current Biology*, *20*, 1573–1579.

Green, A., & Wright, M. (2003). Reduction of task-relevant information in skill acquisition. *European Journal of Cognitive Psychology*, *15*, 267–290.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*, 304–337.

Haider, H., & Frensch, P. A. (1999a). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 172–190.

Haider, H., & Frensch, P. A. (1999b). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied*, *5*, 129–151.

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 392–406.

Haider, H., & Frensch, P. A. (2005). The generation of conscious awareness in an incidental learning situation. *Psychological Research*, *69*, 399–411.

Haider, H., Frensch, P. A., & Joram, D. (2005). Are strategy shifts caused by data-driven processes or by voluntary processes? *Consciousness and Cognition*, *14*, 495–519.

Harris, H. D., Murphy, G. L., & Rehder, B. (2008). Prior knowledge and exemplar frequency. *Memory & Cognition*, *36*, 1335–1350.

Jiang, Y., & Song, J. (2005). Hyperspecificity in visual implicit learning: Learning of spatial layout is contingent on item identity. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1439–1448.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion – Is it all in your head?: Implicit theories about willpower affect self-regulation. *Psychological Science*, 21, 1686–1693.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 830–845.

Lee, F. J., & Anderson, J. R. (2001). Does learning a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology*, *42*, 267–316.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 883–914.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376–400.

Marewski, J. N., & Link, D. (2014). Strategy selection: An introduction to the modeling challenge. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 39–59.

Marewski, J. N., & Olsson, H. (2009). Beyond the null ritual: Formal modeling of psychological processes. *Zeitschrift für Psychologie/Journal of Psychology*, *217*, 49–60.

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, *118*, 393–437.

Myung, J., Kim, C., & Pitt, M. A. (2000). Toward an explanation of the power-law artifact: Insight from response surface analysis. *Memory & Cognition*, *28*, 832–840.

Niessen, C., Eyferth, K., & Bierwagen, T. (1999). Modeling cognitive processes of experienced air traffic controllers. *Ergonomics*, *42*, 1507–1520.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, *55*, 169–195.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method for selecting among computational models for cognition. *Psychological Review*, *109*, 472–491.

Reason, J. (1990). *Human error*. New York: Cambridge University Press.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41.

Rickard, T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 65–82.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*, 258–275.

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science*, *16*, 1–40.

Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 318–341.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, *25*, 203–244.

Touron, D. R. (2006). Are item-level strategy shifts abrupt and collective? Age differences in cognitive skill acquisition. *Psychonomic Bulletin & Review*, *13*, 781–786.

Touron, D. R., & Hertzog, C. (2004a). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging*, *19*, 452–466.

Touron, D. R., & Hertzog, C. (2004b). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition*, *32*, 298–310.

Underwood, G., Crundall, D., & Chapman, P. (2002). Selective searching while driving: The role of

experience in hazard detection and general surveillance. *Ergonomics*, *45*, 1–12.
Vu, K.-P. L., Garcia, F. P., Nelson, D., Sulaitis, J., Creekmur, B., Chambers, V., & Proctor, R. W. (2007). Examining user privacy practices while shopping online: What are users looking for? In M. J. Smith & G. Salvendy (Eds.), *human interface* (Part II, HCII 2007, pp. 792–801). Lecture Notes in Computer Science 4558. Berlin: Springer.
Wenke, D., Gaschler, R., & Nattkemper, D. (2007). Instruction-induced feature binding. *Psychological Research*, *71*, 92–106.
Wagenmakers, E.-J. (2003). How many parameters does it take to fit an elephant? *Journal of Mathematical Psychology*, *47*, 580–586.
Watson, D. G., Maylor, E. A., & Bruce, L. A. M. (2005). Effects of age on searching for and enumerating targets that cannot be detected efficiently. *Quarterly Journal of Experimental Psychology*, *58*, 1119–1142.
Wilkins, N. J., & Rawson, K. A. (2010). Loss of cognitive skill across delays: Constraints for theories of cognitive skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1134–1149.
Woltz, D. J., Gardner, M. K., & Bell, B. G. (2000). Negative transfer errors in sequential cognitive skills: Strong-but-wrong sequence application. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 601–625.

APPENDIX A

Manipulation checks and transfer errors

Manipulation check—Was representation strength successfully varied?

Experiment 1

Three different measures indicated that the frequency manipulation of the letter strings had successfully influenced the representation strength of the strings. First, we computed the median reaction times for correctly verified valid strings separately for each participant, frequency condition, and trial block using Blocks 1–4. As can be seen in Figure A1, reaction times to frequent strings were faster than reaction times to infrequent strings. The slopes of the lines connecting short, medium, and long strings are approximately parallel in each block for frequent, infrequent, and singleton strings. This indicates that the time taken to process irrelevant information does not differ by item frequency. At the same time, the general level of reaction time drops from >4000 ms to around 2000 ms. A 2 (frequency of presentation: three presentations versus one presentation) × 4 (block: 1–4) repeated measures analysis of variance (ANOVA) yielded main effects of frequency of presentation, $F(1,\ 15) = 40.44$, $MSE = 380{,}591$, $p = .001$, $\eta^2_p = .73$, and of block, $F(3, 45) = 53.78$, $MSE = 589{,}001$, $p = .001$, $\eta^2_p = .78$. There was no interaction of frequency of presentation and block ($F < 1$).

Second, participants fixated on the relevant string part of frequent (medium and long) strings less often ($M = 1.74$ fixations) than on that of infrequent or singleton strings ($M = 2.1$ for both of the latter), $t(15) = 3.81$, $p = .002$, $\eta^2 = .49$ for frequent versus infrequent; $t(15) = 4.5$, $p = .001$, $\eta^2 = .58$ for frequent versus singleton.

Third, trials with less frequent strings yielded stronger pupil dilation. This was expected as such trials more often required effortful algorithmic processing of the letter–digit–letter triplet. Specifically, for each trial and person, we calculated a baseline for change of pupil size by taking the median pupil diameter of the 1-s blank period prior to stimulus onset (excluding the 25% data points with the highest rate of change in the fixation coordinates). This baseline was subtracted from the median pupil size of the fixations on the relevant part of the strings (again excluding 25% data points with the highest rate of change in fixation coordinates). Pupil dilation was smaller for the frequent strings ($M = 0.065$ mm) than for the infrequent strings ($M = 0.086$ mm), $t(15) = 2.18$, $p = .045$, $\eta^2 = .24$.

Experiment 2

As expected, frequency of letter presentation affected RT. Means of median reaction times for frequent ($M = 792.3$ ms) were faster than those for infrequent ($M = 968.1$ ms) and one-block-only letters ($M = 974.9$ ms). A 3 (frequency of presentation: frequent, infrequent, one-block-only) × 14 (block: 1–14) repeated measures analysis of variance (ANOVA) yielded main effects of frequency of presentation, $F(2,\ 40) = 47.6$, $MSE = 70{,}124$, $p = .001$, $\eta^2_p = .7$, and of block, $F(13,\ 260) = 40.03$, $MSE = 41{,}023$, $p = .001$, $\eta^2_p = .68$. There was no interaction of frequency of presentation and block ($F = 1.3$).

Transfer errors—When the formerly irrelevant part becomes relevant

Experiment 1

Transfer error rates were high (Figure A2a). Eleven out of 16 participants had a transfer error rate higher than 90%. Only the (scarce) singleton data suggested an effect of frequency of presentation on transfer errors, otherwise $F < 1$. Consistent with the once-for-all view of item-general information reduction, participants failed to detect violations of the alphabetical order in the formerly irrelevant part of frequent, infrequent, and singleton strings alike.

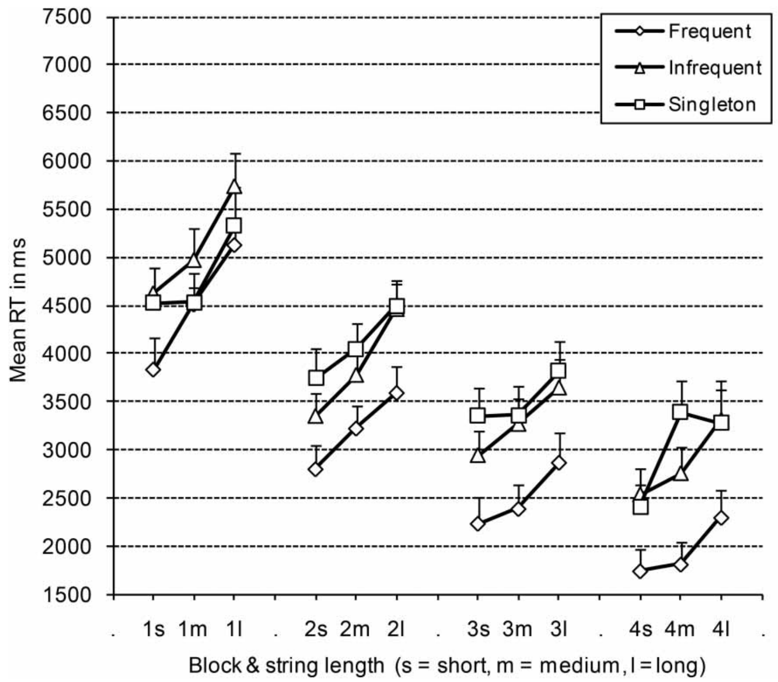

Figure A1. *Means of individual median reaction times (RTs) for Experiment 1. On the* x*-axis, reaction times for short, medium, and long strings (s, m, l) are grouped together by block in order to display the processing amount of irrelevant information.*

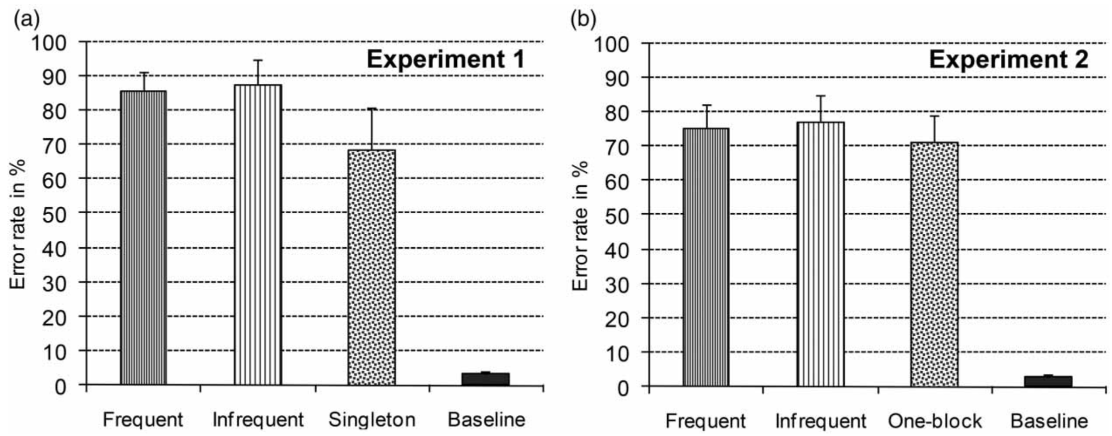

Figure A2. *Transfer error data from the test block of Experiment 1 (Panel a) and Experiment 2 (Panel b) are displayed, together with the mean error rate on regular trials in the same block as baseline.*

Experiment 2

Again, transfer error rates were high. Twelve participants had a transfer error rate $\geq 90\%$. There was no effect of frequency of presentation on the rate of transfer errors, $F = 1.1$ (Figure A2b).

Dynamics of fixations within trials—Information reduction includes first fixations

For Figure A3 we calculated the proportion of first, second, … eighth fixations within trials falling on a relevant versus an

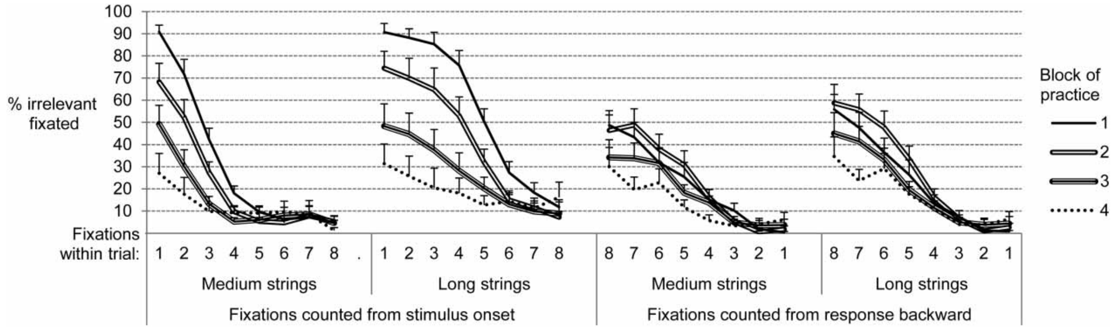

Figure A3. *Percentage of first to eighth fixations within trials that fall on irrelevant aspects of medium and long strings are displayed per block of practice. Fixations are counted either from stimulus onset (increasing numbers) or backwards from the response.*

irrelevant letter position. We did so by counting forward from stimulus onset and in addition ran the equivalent analysis by counting backward from the response (1 = last fixation prior response, 2 = prelast fixation prior response, … 8 = last fixation prior response). Figure A3 does not provide evidence that information reduction might start with medium-length strings and later generalizing to long strings. We observed that in Block 1, first fixations almost exclusively fell on irrelevant parts of alphanumeric strings. Later fixations fell on relevant character positions. In later blocks of practice, also a large proportion of first fixations fell on relevant string positions. Apparently this was the case for medium and long strings as well. There was no indication that late fixations within trials might be devoted at (finally) checking the irrelevant letter positions (which early in practice had been checked first within a trial).

APPENDIX B

Model mimicry analysis

When selecting between competing models, *model mimicry* can become an issue (e.g., Myung, Kim, & Pitt, 2000). Model mimicry refers to a model's ability to fit not only data generated by its own process, but also data generated by some other model. For instance, Haider and Frensch (2002) showed that the average of multiple step functions nicely fits the power function if the probability of a shift point occurring follows a power function. Here we tested the reverse concern—namely that the step function might even fit well to data that originate from a process that is described best by a continuous learning curve, specifically the power function. However, as it turns out, model mimicry was not a tenable account of the good fits of the step function. We compared the distribution of change trials derived from fitting step functions to either data or estimated values of the alternative model: (a) the fixation data and (b) the estimated values of power function fits of the four frequent and the four infrequent items of each participant. Sorted within frequency class and pooling over participants, Figure B1 depicts where, in terms of change trials, the 64 steps are located (16 participants with four frequent and four infrequent items). The step function fits derived from the data had change trials that were evenly distributed over the whole experiment. However, the change trials derived based on the power function model were located within the first 150 trials.

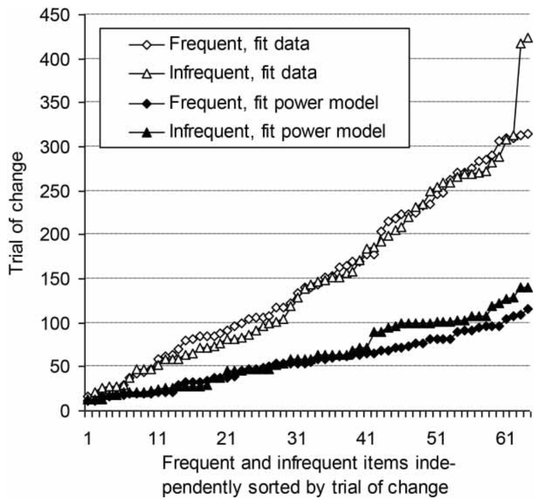

Figure B1. *Trials of change (y-axis) across practice are depicted as derived with three-parameter step function fits for the four frequent and four infrequent items per participant together with step function fits based on power function approximations of the data. Frequent and infrequent items for 16 participants total 64 items (x-axis), sorted independently in ascending order according to trial of change.*