Memory & Cognition (2019) 47:696–705
https://doi.org/10.3758/s13421-018-0881-x

Learning how to exploit sources of information

Brad Wyble[1] · Michael Hess[2] · Ryan E. O'Donnell[1] · Hui Chen[3] · Baruch Eitam[4]

Abstract

How is our strategy for forming memories shaped by experience with a task? Previous work using surprise questions (i.e., unexpected by the participant) has shown a remarkable inability to report attributes of an attended target in a search display. This representational poverty presumably reflects a form of *information exploitation*, in which control processes specialize the conversion of available information into memory representations. We hypothesize that such control is refined by repeated experience with a task, and as a result, memory representations will specialize as task experience accrues, such that report accuracy for an unexpected question will progressively worsen as the number of preceding trials increases. To test this, subjects were asked to report the location of a letter among three digits. The ability to respond correctly to a surprise question about the identity of that letter became worse as the experiment progressed. A follow-up study evaluated whether this incremental worsening of report accuracy could be explained as a buildup of proactive interference by varying the set of letters for the surprise test. The results were unchanged relative to the original experiment, which argues against a primary contribution of proactive interference in the worsening performance. The effect was replicated in a similar paradigm using color disks. These findings illustrate that repeated performance of a prescriptive task engages an adaptive modification of control processes that focus information processing on specific attributes of a stimulus that are expected to be necessary in the future, regardless of their immediate task relevance.

Keywords Working memory · Control processes · Attribute amnesia · Exploration/exploitation · Irrelevance induced blindness

It has often been assumed that visual attention and working memory are tightly intertwined (Awh & Jonides, 2001; Downing, 2000; Theeuwes, Belopolsky, & Olivers, 2009), and therefore one might predict that attending to an object should cause information associated with that object to be stored in memory. However, recent findings have shown that our ability to report information from even simple stimuli is quite selective, such that seemingly task-relevant attributes of an attended object can be inaccessible to report in a probe presented immediately after the stimulus. For example, when participants are asked to locate a target in an array according to a specific attribute (e.g., find a letter among digits, or find a colored letter among black letters) for 10 or more trials in a row, they can easily report the target's location but do not reliably report other attributes of that object regardless of task relevance. This failure to report a seemingly obvious and task-relevant target attribute is obtained with a surprise question (e.g., report the identity of the letter) posed directly after the offset of the search array. For some of these experiments, accuracy is indistinguishable from chance in a four-choice recognition test (H. Chen & Wyble, 2015a; Chen & Wyble, 2016; Swan, Wyble, & Chen, 2017). On the very next trial when that same surprise question is asked (now no longer a surprise), performance becomes dramatically better. This improved performance persists for a series of control trials, showing that participants are easily able to report the target's defining attribute once they have been exposed to the surprise question.

This inability to report the target-defining attribute in a surprise question seemingly reflects the operation of a type of control process described by Atkinson and Shiffrin (1968). Information (e.g., about the identity of the target letter) that has passed through the sensory register and been attended to some degree, is not entered into a reportable representation, presumably because, according to both task instructions and the previous experience of the participant, this information is no longer necessary. The phenomenon of being unable to

---

✉ Brad Wyble
bwyble@gmail.com

[1] Psychology Department, Pennsylvania State University, University Park, PA, USA

[2] Donders Institute, Radboud University, Nijmegen, The Netherlands

[3] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China

[4] Psychology Department, University of Haifa, Haifa, Israel

report this key attribute of a target, despite the fact that the same attribute had relevance for the task of detecting the target mere moments before, reflects a form of competence in *information exploitation*, a term that we elaborate below.

In a similar series of findings, it has been shown that this impairment in reporting an attribute of a target object that is unexpected happens even on the *first trial* of an experiment (Eitam, Yeshurun, & Hassan, 2013), which indicates that the expectation to remember can be induced by task instruction alone. This was termed *irrelevance blindness* and refers to the idea that information thought to be irrelevant to a task can be hard to remember even if it was near the focus of attention. However, performance on this first-trial question was substantially better than it was in H. Chen and Wyble (2015a), a case in which the surprise trial occurred after a substantial number of trials. This difference presumably indicates that experience with the task enables participants to develop an entrenched task set that allows them to locate the target without processing its identity in a way that allows them to remember it (see also Eitam, Shoval, & Yeshurun, 2015).

However, the two tasks differ on various dimensions; the task used by Eitam et al. (2013) involves the unmasked presentation of a single stimulus, for a relatively long duration (500 ms), and the probed attribute is completely task irrelevant, while the task used by the H. Chen and Wyble (2015a) involves the presentation of four objects for 150 ms and the probed attribute (i.e., target-defining attribute) has been relevant to locating the target. The goal of this paper is to bridge the gap between these results by showing how task experience affects what information can be reported about a stimulus.

Theoretical premise

The theoretical premise underlying this investigation is that control processes act to minimize the amount of interference in the cognitive system, particularly when an ongoing task is highly prescriptive. The theory takes account of the fact that activating any piece of information in memory causes interference for other information, either at encoding (e.g., Eich, 1985; Murdock, 1982; Swan & Wyble, 2014) or retrieval (Anderson & Bower, 1972; Gillund & Shiffrin, 1984; Murnane & Shiffrin, 1991). Therefore, it should be advantageous to store or even activate only particular pieces of information in memory, based on expectations of future relevance during performance of the task.

On the other hand, when the longer-term task demands are new or uncertain, a greater proportion of available information will be stored or activated on the chance that such information will become relevant at some later point in time.

This distinction is analogous to the classic distinction between exploration and exploitation (e.g., March, 1991; Mehlhorn et al., 2015), in which an agent must decide between consuming resources from one known source (i.e., exploiting) or searching for potentially higher valued sources (i.e., exploring). In this case, the distinction is applied to information rather than more tangible resources or locations. In this framing, when task demands are uncertain, memory functions in a more liberal fashion by encoding or activating a larger proportion of potential information from the environment, defined here as *information exploration*. The analogy with exploration is due to the fact that the extra information stored has uncertain value and comes with a cost due to increasing interference with other memories. As requisites of a given task become sharper due to experience, the system shifts into a more exploitive mode of processing that restricts which information will be accessible in the future, a term which we referred to previously as *information exploitation*.

The question addressed in these experiments is whether we can experimentally observe the systematic switching from an exploratory to an exploitive form of information processing over the course of multiple trials. This will be done with two similar approaches, one being the case of attribute amnesia (H. Chen & Wyble 2015a) and the other irrelevance blindness (Eitam et al., 2015).

Experiment 1

In Experiment 1, we address this question by sorting participants into five groups, each of which replicates the canonical attribute amnesia paradigm of H. Chen and Wyble (2015a), but with variation in the timing of the surprise trial relative to the start of the experiment. The central prediction is that performance in the surprise trial will progressively worsen as participants become more experienced with the task, since they develop a more exploitive version of the task set. Note that each participant will experience the surprise question only once, meaning that different groups of participants are necessary to probe the state of memory encoding at different trial points.

Method

The scripts for executing the experiments, the data files, and the analysis code are posted here:

https://osf.io/5g9nh/

The preregistration for the crucial analyses are listed here:

https://osf.io/f3jhm/register/5771ca429ad5a1020de2872e

Participants The experiment was approved by the Penn State University Institutional Review Board. One hundred undergraduate participants were included after exclusion according to the specified criteria in our preregistration that performance had to be above 60% in reporting location, color, and letter identity during the last five trials of the experiment. This criterion ensured that subjects could perform the more complex control trials reasonably well, and these particular trials were chosen in our preregistration because we wanted to see how performance rebounded on the control trials immediately after the surprise trial. This excluded five participants. The participants were divided pseudorandomly into five groups that determined at what point during the trial sequence they experienced the surprise trial (Trial # 1, 2, 5, 10, or 50). The sample size of 20 per group was chosen to match earlier publications of this same paradigm (e.g., H. Chen & Wyble, 2015a).

Apparatus The experiment was run on MATLAB 2012, using Psychtoolbox (Version 3; Kleiner et al., 2007) on Windows XP. The display was a CRT monitor with a refresh rate of 75 Hz at 1024 × 768 resolution.

Procedures At the start of the trial, participants were given the following instructions:

> “On each trial you will see one colored letter (target) and three colored digits (distractors) flashed on the screen. All of these four stimuli will be masked very quickly. Try to locate the ‘colored letter’ because you need to report the location of the colored letter at the end of the trial! The digits are irrelevant to your task, so just ignore them.
> If you are ready, please press the Spacebar to start.”

Refer to Fig. 1 for a visual depiction of the experiment. The display background was gray (RGB [150, 150, 150]). On each trial, participants were shown a display consisting of a central fixation cross and four circular placeholders, 0.52° across, arranged in a square of width 5.5°. After a randomized latency of 800–1,800 ms, the four placeholders were replaced by three digits (2–9 selected without replacement) and one letter (A, B, D, or E). The color of each stimulus was one of red (RGB [200, 0, 0]), blue ([0, 0, 200]), yellow ([200, 200, 0]), or purple ([190, 45, 200]) selected without replacement. After 159 ms, the array was replaced by four masks composed of a black ‘@’ symbol, and four lines composed of the four stimulus colors, always in the same configuration. After 106 ms, the display was blanked for a duration of 400 ms. Then followed one of two question sequences depending on the condition and trial number.

Presurprise trials Early trials were designated as presurprise trials, in which participants were asked the question they expected based on the instructions, which was to “Press a corresponding number to indicate the location of the colored letter.” At the same time, the digits 1, 2, 3, and 4 were shown in the same location as the items in the array, and participants had to choose one, after which they were given feedback as to the target’s correct location.

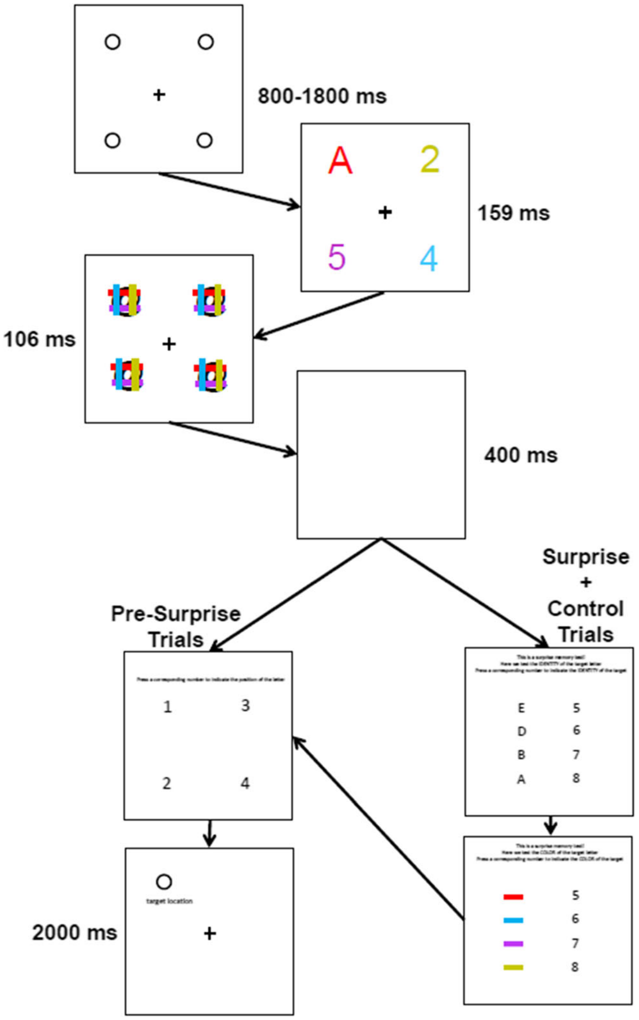

Fig. 1 Paradigm for Experiments 1 and 2. The top four frames occurred for all trials. On presurprise trials, participants only saw the location question, while on the surprise and control trials, they saw the questions about identity and color. (Color figure online)

Surprise and control trials The initial portion of the surprise trial was identical to the presurprise trial, including the placeholders, the stimulus array and the retention interval. Instead of the location question, participants were asked to report the identity of the letter using a four-alternative forced-choice question that read:

> “This is a surprise memory test!
> Here we test the IDENTITY of the letter.
> Press a corresponding number to indicate the IDENTITY of the target letter.”

Below this question was a four-alternative forced-choice as shown in Fig. 1. The ordering of the four letters was randomized on each trial. Then, participants were asked to report the letter's color using a similar question, and finally they were asked to report its location using the question format they had previously been exposed to. The surprise question was followed by 15 identical trials (i.e., control trials), each of which asked the same three questions in the same order.

The position of the surprise trial in the sequence was determined by their condition, such that "Surprise-*N*" participants performed the location task in each of $N - 1$ presurprise trials before the onset of the surprise trial. For example, Surprise-1 participants were asked the surprise question immediately on the first trial, Surprise-2 participants were surprised on the second trial after they had experienced the location question on the first trial, and so on for groups Surprise-5, Surprise-10, and Surprise-50.

The total number of trials experienced by each participant varied across the five conditions, being 16 for the Surprise-1 participants (the surprise trial + 15 control trials), 17 for Surprise-2, 20 for Surprise-5, 25 for Surprise-10, and 65 for Surprise-50.

The primary analysis of interest was a permutation test on the slope of the accuracy scores across the five conditions to see if accuracy changed in a systematic fashion using MATLAB's polyfit function. To determine significance using a permutation analysis, this observed slope was compared to a null distribution of slope values computed by randomly shuffling the 100 participant condition labels across the five conditions, 10,000 times.

As a measure of how participants reacted to the surprise question, an additional analysis, which was also mentioned in the preregistration, computed the slope of the median RT to the first surprise question (letter identity) across the five conditions in a similar fashion. Median was chosen as a measure of central tendency to offset the skew of the RT distributions.

Results

The primary behavioral results are listed in Table 1. Note that there is a variable number of trials across cells. For example, for the condition "Surprise-2" there is only one presurprise location trial. The data from the control trials include the mean of the 10 trials immediately following the surprise trial and excludes the last five trials, since those were used for participant exclusion.

Preregistered analyses The key analyses explored the pattern of performance on the surprise questions about the identity and color of the target letter that had just been located. The prediction was that performance would get steadily worse as the surprise trial occurred progressively later in the trial sequence. The analysis used a permutation analysis that computed the slope of accuracy scores across the five conditions as described in the Method section.

For report of the target identity, the relationship between accuracy and trial number was significantly different from the null distribution in a one-tailed analysis, slope = −.075, $p < .02$. However, for color report, the slope was not significantly different from the null distribution, $p > .15$, although performance was so close to chance that there was little room for performance to get worse.

To examine the time course of changes in the task set following the surprise question, Table 2 compares accuracy for each of the three questions between the surprise trial (data copied from Table 1) and the immediately following trial (i.e., the first control trial). This illustrates how rapidly participants can alter what information should be accessed/stored. For example, in the most extreme example (Surprise-50 group), participants improved from 30% to 90% for the letter identity question, which demonstrates that it was not difficult to report these additional attributes when prepared to do so. Moreover, the rapid change illustrates that the changes in processing strategy due to task entrenchment can be rapidly modified.

Reaction time The preregistration specified a secondary analysis of reaction times to the surprise question to determine how experience with the task affects reaction time when participants first encounter the surprise trial. There is comparatively little data on reaction times to unexpected events so there was no a priori prediction of what would be observed. Moreover, these are not speeded responses as used in typical RT paradigms. The opportunity to respond is only available 506 ms after the stimulus array was presented, and participants have to read a screen full of text prior to responding when each

Table 1 Percentage accuracy for a variety of questions at different points in the experiment

Accuracy	Presurprise location	Surprise identity	Surprise color	Surprise trial location	Control identity	Control color	Control location
Surprise-1	N/A	60	30	90	84.0	82.5	82.5
Surprise-2	80	55	35	85	85.0	81.5	84.5
Surprise-5	72.5	45	30	75	87.0	89.5	87.0
Surprise-10	80.0	40	30	80	86.0	84.5	85.0
Surprise-50	81.6	30	25	80	89.0	88.0	88.0

Table 2 Comparison of surprise trial performance to the immediately following control trial. Numbers represent percentage of accuracy. The key takeaway is the decline in accuracy of the surprise identity question

Accuracy	Surprise identity	First control identity	Surprise color	First control color	Surprise trial location	First control location
Surprise-1	60	85	30	75	90	70
Surprise-2	55	95	35	90	85	90
Surprise-5	45	100	30	100	75	100
Surprise-10	40	85	30	70	80	70
Surprise-50	30	90	25	85	80	85

surprise question is presented. Despite these limitations, the data may be indicative of cognitive processes elicited by the surprise question.

The exploratory analysis, using a similar permutation approach as above, showed a highly significant change in median reaction time (mRT) with increasing task experience ($p < .0001$). The same analysis conducted on the mRT to the color surprise question was not significant ($p > .06$) and there was even less variability in the mRT to the location question (Table 3).

Discussion

The pattern of accuracy scores for the surprise trials suggests that near the beginning of the trial sequence, when participants had only the instructions and little experience with the task, they accessed the target letter’s identity along with its location and stored these values in memory. However, with continued task experience, participants formed progressively weaker representations of the letter’s identity such that they were nearly at chance performance by the time they had completed 50 trials. The implication is that over the course of the experiment, as participants accrued more task experience, they adopted a more rigid task set that allowed them to locate the target without encoding its identity. In contrast, there was very little evidence of an accessible memory for the target letter’s color in any of the conditions. Forgetting may have contributed to this poor performance on color report, but participants were quite accurate in reporting the location information, which was the third question on the surprise trial.

An alternative explanation for the gradual decrease in surprise trial accuracy is proactive interference, since the same four letters are used repeatedly across the entire trial sequence. Thus, the ability to report the letter's identity may be masked by numerous memories of the letter from previous trials rather than by the changes in memory control processes. This possibility will be addressed in Experiment 2.

Experiment 2

To address the possibility that the decline in accuracy was due to proactive interference, an additional experiment replicated the Surprise-50 condition from Experiment 1, except that the letter set was changed prior to the surprise trial. In this experiment, the set of four potential target letters changed at trial 47, so that the surprise trial, and the three trials before it, each used a different letter. If the accumulation of proactive interference was the cause of the decline in surprise trial accuracy across the five conditions, then this manipulation should bring release from it and thus increase accuracy on the surprise trial dramatically. In short, it did not.

Table 3 Median reaction time (mRT) to particular questions at different points in the task. Reaction time is measured relative to the onset of each question, the earliest of which occurred at least 500 ms after the stimulus onset, and reported in seconds. The key takeaway is the increase in mRT for the surprise question across conditions

Median Reaction Time (mRT)	mRT to surprise letter	mRT to surprise color	mRT to surprise trial location	mRT to First control letter	mRT to First control color	mRT to First control location
Surprise-1	15.2	6.0	5.4	5.3	2.2	2.1
Surprise-2	15.1	5.7	5.7	6.6	2.8	2.8
Surprise-5	21.8	8.9	5.4	6.5	2.9	2.8
Surprise-10	17.7	8.0	5.9	4.4	2.1	2.0
Surprise-50	24.0	7.0	5.7	5.4	1.6	1.5

Method

The scripts for executing the experiments, the data files, and the analysis code are posted here:

https://osf.io/q9egf/

The preregistration for the crucial analysis is listed here:

https://osf.io/6eyzk/register/5771ca429ad5a1020de2872e

Participants A new sample of 26 participants was obtained, and the first 20 participants that passed the trial inclusion criteria were selected. Due to poorer-than-expected performance by participants, we modified our participant exclusion criterion to be less restrictive to ensure that at least 20 participants passed. Rather than 60% accuracy on location, color, and letter identity, we relaxed the restriction on color accuracy to 40% with the rationale that color accuracy is unimportant to the analysis being conducted here. This decision was made before viewing the key dependent variable, which was accuracy on the surprise trial for letter identity.

Procedure All participants were included in a single group that received the surprise trial on Trial 50. The only change was that the target letter set changed without warning from ABDE to GMRT on Trial 47. These letters were chosen so as to be clearly distinct. Furthermore, on Trials 47 through 50, the four letters were randomly shuffled, instead of being randomly chosen to ensure that there were no repeats within that first set of four trials using the new letters. On the remaining control trials target letters were selected at random from the new set of letters. The only other change was that the letter identity response display used only the new set of target letters for the surprise and all control trials.

Results

The results from this experiment are shown alongside the comparable condition from Experiment 1 (see Table 4). The preregistered analysis would have compared the accuracy in the three questions during the surprise trial across the two experiments using a chi-square in search of evidence that the manipulation improved accuracy on the surprise trial. This comparison is unnecessary for the identity and surprise trials in the event since performance was near chance on both questions in both experiments. For location, the difference was not significant ($\chi^2 = 2.85, p > .09$). The lower performance on the location trial may have to do with variability in forgetting, since this was the third question asked. In previous work using this paradigm (H. Chen & Wyble, 2015a), location accuracy on the surprise trial varied considerably across experiments. Most importantly, the accuracy of identity report on the surprise trial of this experiment was identical to that in Experiment 1 (30% correct in both experiments), indicating that the introduction of a new set of letters just before the surprise letter identity question made no difference to the performance on the identity question. This finding discredits the possibility that accumulated interference is primarily responsible for the monotonic decrease in the ability to report on the identity of the letters. Note that these results are in contrast to the findings of W. Chen and Howe (2017), who presented a modified version of this paradigm that used small images of objects as stimuli and asks participants to report the location of the animal. They found that when the image was novel on each trial, participants could accurately remember the target on a surprise trial, even though the surprise trial was preceded by 155 location-question trials. This difference indicates an improvement in retrievability due to the larger variability in featural and semantic aspects of the set of animal images compared with the highly familiar capital letters used here.

Experiment 3

An experiment was run to test whether the onset of exploitation would be obtained using the irrelevance blindness effect (Eitam et al., 2013). This experiment was actually run prior to the previous two experiments in an attempt to understand why performance differed in the Eitam et al. (2013) and H. Chen and Wyble (2015a) experiments. As such, it was not preregistered but is included as it provides additional evidence of the generality of the observed effect. The experimental protocol was designed to replicate as precisely as possible the paradigm of Eitam et al. (2013).

Table 4 Comparison of Surprise-50 groups between experiments. In Experiment 1, the letters ABD and E were used for all of the trials. In Experiment 2, letters GMR and T were introduced without repetition on trials 47 through 50, and in the subsequent control trials with repetition

Accuracy	Presurprise location	Surprise identity	Surprise color	Surprise trial location	Control identity	Control color	Control location
Surprise-50 Exp 1	81.6	30	25	80	89.0	88.0	88.0
Surprise-50 Exp 2	81.0	30	25	55	85.0	78.0	80.0

Method

The scripts for executing the experiments, the data files, and the analysis code are posted here:

https://osf.io/jfmru/

One hundred and one participants were shown a pair of concentric circles (inner diameter 1.4°, outer diameter 2.8°). As before, there were five groups, each receiving a surprise trial at a different point in the experiment. While the goal was 100 participants, several data files were invalid. Also, the Surprise-10 group received their surprise trial on trial 20 due to a coding error and will be termed Surprise-20. The counts of participants per group were (Surprise-1: 18, Suprise-2: 20, Surprise 5: 19, Surprise-20: 20, Surprise 50: 19). See Fig. 2 for a visualization of the task.

The colors of the circles were chosen without repetition from purple (RGB [128, 0, 128]), orange ([255, 153, 0]), red ([255, 0, 0]), pink ([255, 105, 226]), green ([0, 176, 80]), blue ([0, 112, 192]), and yellow ([255, 255, 0]). The participant always viewed the stimulus for 500 ms and then a blank screen for 500 ms before being asked to report a color.

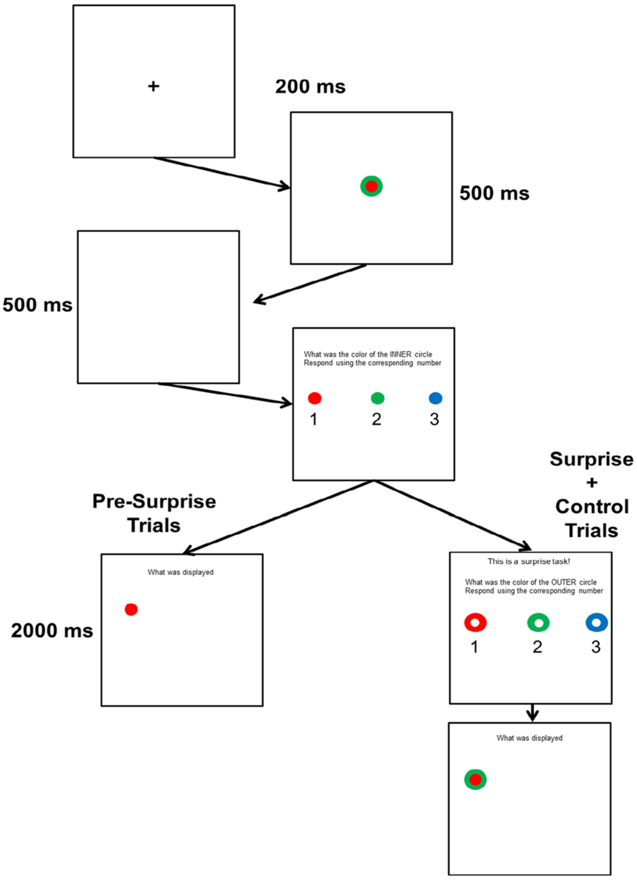

Fig. 2 Paradigm for Experiment 3. The top four frames occurred for all trials, and the bottom frames occurred selectively depending on trial number and participant group. (Color figure online)

On the presurprise questions, half of the participants were asked, "What was the color of the INNER circle?"; the other half were asked, "What was the color of the OUTER circle?" The question was accompanied by three colored shapes, in the same size and shape as the original portion of the stimulus that was being queried (a circle or a donut). One of the three colors corresponded to the correct answer, and the other two were chosen randomly from the remaining six.

On the surprise trial, participants were asked the same question first, and then asked the other question (i.e., they were probed about the color of the circle that they would have assumed was irrelevant after being asked about the relevant question) (Table 5).

In a similar permutation test as was done for Experiment 1, the slope of performance for Surprise Trial 2 was computed across the five conditions. This slope was compared against a null distribution of permuted condition labels as before. Out of 10,000 permutations, the permuted slopes were more extreme than the actual slope in only two cases, providing a p of < .001. Thus, a similar transition from exploration to exploitation occurs in the irrelevance blindness effect, similar to the pattern of results obtained in the attribute amnesia experiments.

Unlike in Experiment 1, reaction times to the second question did not follow the same monotonic increasing trajectory and were much faster, despite the fact that the question was unexpected. This is presumably a result of the similarity between the form of the surprise question and the presurprise question, both of which required participants to report a color using a number key press.

General discussion

These results measure the time course of the onset of information exploitation with experience in a task. At the beginning of the experiment, when participants had to detect a target according to its visual form (i.e., find a letter among numbers), they were typically able to report the identity of that letter in response to an unexpected "surprise" probe. However, with more experience on the task prior to the surprise probe, participants had progressively poorer accuracy in reporting the letter's identity. A similar effect can be observed when subjects are asked to report a completely irrelevant stimulus in a surprise trial, such that performance becomes progressively worse as the surprise question occurs later in the trial sequence. The similarity between the two results suggests that the letter's identity in Experiment 1 was treated by memory encoding processes in a similar fashion as the completely irrelevant color in Experiment 3, despite the fact that this attribute had some relevance for selecting the target in the first place. This raises a question as to the definition of task relevance, which typically refers to the relevance to the entire task.

Table 5 Accuracy for the three critical question types used in the replication of Eitam, Yeshurun, and Hassan (2013). Color-1 refers to the first question that was asked (either about the inner or outer color for each participant) and was always expected. Color-2 was a question about the other color and was unexpected on the surprise trial. Reaction times are in seconds

Accuracy	Color-1	Surprise trial Color-1	Surprise trial Color-2	mRT to Color-2
Surprise-1	NA	88.9	94.4	4.2
Surprise-2	95	100	55.0	4.3
Surprise-5	97.4	100	52.6	4.3
Surprise-20	98.4	100	40	6.2
Surprise-50	98.9	100	36.8	5.4

However in this case, the identity of the letter in Experiments 1 and 2 can be said to be irrelevant to the memory requirements of the task.

This effect can be construed as a form of information exploitation, such that experience with a task gives rise to a highly specific task set which "exploits" available information resources by either (1) only encoding those pieces of information that will be necessary in the future, regardless of immediate relevance, or (2) categorizing stimuli at the most efficient level for performing the task thus minimizing the activation of irrelevant information (Eitam & Higgins, 2010)—here shifting from a stimulus' identity to its category (letter/digit). What is notable about this demonstration is that task relevance alone is not fine-grained enough to predict "downstream" processing, since, the identity of the stimulus was the target-defining attribute. Thus, the results demonstrate the precision of control processes to operate on attributes that are integral to the target in a search display.

Exploitation can be contrasted with the more exploratory mode of information processing presumed to operate at the beginning of a block of trials, in which participants have been instructed how to perform the task but do not yet have experience performing it. It is also an interesting observation on the limitations of verbal instruction in fine tuning cognitive control processes.

Notably, there is little evidence that the target's color, which is entirely task irrelevant from the participants' perspective, creates a reportable impression in memory at the start of the experiment. Even on the very first trial, performance on the color surprise question is near the level of random guessing. This suggests that the task set established by the instructions is sufficient to reduce memory encoding of completely irrelevant information, a result that is in agreement with the findings of Eitam et al. (2013), although the effect is stronger here. The worse performance in this case is likely due to additional factors, such as the fact that four stimuli rather than one were shown in the key display, a mask was used, and color was the second question asked during the surprise trial. Note that even though in Eitam et al. (2013) the participants' set was fully determined by a location's relevance, we have evidence that the same adaptation of control processes (or "exploitation mode") occurs in that paradigm, too, in Experiment 3.

Like the Mack and Rock (1998) studies on inattentional blindness, the surprise trial is followed by a series of control trials to illustrate that participants are able to extract all three attributes (location, color and identity) from a target once they have an expectation to do so (see also Eitam et al., 2013, Exp. 2). Participants achieve comparably high levels of accuracy on all three of these reports over the following trials.

The analysis also explored reaction time to the surprise question. RTs to such questions are difficult to interpret because they are quite long (>10 seconds) and encompass a wide variety of cognitive processes. For example, participants must read the text on the screen, attempt to retrieve the information, and then generate a response using the prescribed keys. There is no time pressure placed on the participant so that an RT of 20 seconds might include a considerable amount of rumination as participants attempt to retrieve the information and fail repeatedly. Despite these complications, the surprise questions were exactly the same among groups and there was a clear and highly significant trend of changes in the median reaction time across conditions, such that RTs to the surprise question became longer with increased experience in the task. This finding supports the idea that participants had become more deeply entrenched in a particular way of performing the task and thus required more time to disengage.

Another key finding from the RT analysis is the enormous difference between the first (letter = 15.8 seconds) and the second surprise question (color = 7.12 seconds) on the surprise trial. Given that the second surprise question is similar in complexity to the first surprise question, it is clear that the 15 seconds that participants spend on the first surprise question is not entirely devoted to reading and generating a response. This discrepancy points to the cognitive costs of adjusting to the novelty of being surprised by a question. Furthermore, this cost seems to increase as the experiment progresses, presumably reflecting the difficulty of mentally switching gears out of the exploitation mindset. It would be interesting to unpack this effect of "surprise" into multiple processes in future work. Another interesting avenue is to determine the rate at which different kinds of stimulus attributes become difficult to remember. For example in our previous work, subjects provided accurate report of the location for a spatial cue during a surprise task even after a considerable number of trials (H. Chen & Wyble, 2015b). At the time we concluded that location was automatically encoded, but it is perhaps more likely that different kinds of attributes are differentially resilient to information exploitation.

Mechanisms of information exploitation

As mentioned above, the mechanism of task entrenchment that leads to exploitation could take at least two forms that may be discerned in future work. One form is that the identity of the letter is extracted and fully processed during target detection, but this information is then discarded without being encoded into memory by a control process (Atkinson & Shiffrin 1968), or perhaps encoded so weakly that it is rapidly forgotten during the surprise question (see Jiang, Shupe, Swallow, & Tan, 2016, and subsequent comments by Swallow, Jiang, & Tan, 2017; Swan et al., 2017; Wyble & Chen, 2017). This conclusion was originally suggested by Experiment 2 of H. Chen and Wyble (2015a), wherein it was found that even after using digit parity to locate a target, participants could not report the identity of the digit in the surprise trial. It was assumed, therefore, that participants had to have accessed and then discarded the digit's identity since parity is not typically thought to be a preexisting perceptual category (e.g., Dehaene, Bossini, & Giraux, 1993). In this scenario, there are two mechanisms that might prevent the letter identity from being encoded into memory. First, the identity of the letter, being fully activated during the process of detecting the target, might be actively inhibited to ensure that it does not enter memory. Such a mechanism would be consistent with the claim that it is vital to protect memory resources from being wasted on information that is unlikely to be needed in the future. The second possibility is that a memory filter passively blocks the letter identity from being encoded. This would be consistent with the idea proposed by Botella, Barriopedro, and Suero (2001) that there are distinct filters for determining what information is used to trigger attention (the key feature) and what information is stored for report (the response feature).

Alternatively, exploitation may occur at an even earlier stage of processing. Control processes could allow the digit/letter category of the target to be extracted without activating the representation of its identity at all. Such an explanation would also explain the results of Experiment 2, wherein switching to new letters provides no benefit for target report. Similarly, in Experiment 3, participants may not process the identity of the irrelevant color. If participants were categorizing letters without processing their individual identity (such as "letter" vs. "digit"), then there is no reason that switching the set of target letters would alter performance. It is noteworthy, however, to mention that W. Chen and Howe (2017) found that using a novel target image on each trial largely eliminated the attribute amnesia effect, which suggests that the identity of the stimulus is determined during the process of finding the target. However, since the targets were novel on each trial, it may have been hard to adopt a strategy that allowed categorization without identification. The idea that participants can learn how to categorize stimuli without identifying them first was originally suggested by Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977), in experiments that explored how participants optimize performance to a highly consistent task set.

Thus there are essentially three different avenues by which information exploitation could happen: (1) the information is activated and then actively suppressed, (2) the information is activated and excluded from encoding by a memory filter, or (3) the information is not activated in the first place because a more parsimonious target categorization is developed. The present results are unable to distinguish between these possibilities and the determination is left to future work.

Conclusion

When performing the same task over and over, there is a progressive shift in the processing of information such that the ability to remember attributes of a stimulus becomes focused on those that are expected to be important for the task. This focus has the effect of reducing the ability to report other attributes of a stimulus, even when those attributes are integral to an attended target, are to some degree task relevant, and the unexpected question is posed almost immediately after seeing and attending the target item. In this context, the failure to report in the surprise question reflects the development of an information exploitation strategy that may be essential for restricting the amount of interference generated by processing of stimulus dimensions that are likely to be unnecessary for a given task. Many questions remain to be explored. For example, it is unclear whether a control process that reduces memory encoding of information thought not to be relevant does so by gradually weakening memory traces, or operates in a more dichotomous fashion. Furthermore, how much exposure to a question is required to cause control processes to maintain encoding of a given piece of information? For example, are very rare questions sufficient to keep a given type of information within the set of exploited sources, and how does a contextual change alter these settings? Answering questions of this sort will be helpful for understanding how information encoding strategies change with exposure to new demands.

Acknowledgements This project was completed with the support of a Binational Science Foundation Grant No. 2015299 to B.W. and B.E. This work was also funded by grants from the National Natural Science

Foundation of China (No. 31771201), Humanities and Social Sciences Foundation of the Ministry of Education of China (No.17YJA190001), the Natural Science Foundation for Distinguished Young Scholars of Zhejiang Province (No.LR19C090002) and fundamental research funds for the central universities (2017QNA3018) to author H.C.

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Amsterdam, The Netherlands: Elsevier.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Botella, J., Barriopedro, M., & Suero, M. (2001). A model of the formation of illusory conjunctions in the time domain. *Journal of Experimental Psychology: Human Perception and Performance, 27*(6), 1452–1467.

Chen, H., & Wyble, B. (2015a). Amnesia for object attributes: Failure to report attended information that had just reached conscious awareness. *Psychological Science*, *26*(2) 203–210.

Chen, H., & Wyble, B. (2015b). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research*, *107*, 76–85.

Chen, H., & Wyble, B. (2016). Attribute amnesia reflects a lack of memory consolidation for attended information. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 225–234.

Chen, W., & Howe, P. D. L. (2017). Attribute amnesia is greatly reduced with novel stimuli. *PeerJ*, *5*, e4016.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122(3),* 371–396.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473.

Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 764–770.

Eitam, B., & Higgins, E. T. (2010). Motivation in mental accessibility: Relevance of a representation (ROAR) as a new framework. *Social and Personality Psychology Compass, 4*(10)*,* 951–967.

Eitam, B., Shoval, R., & Yeshurun, Y. (2015). Seeing without knowing: Task relevance dissociates between visual awareness and recognition. *Annals of the New York Academy of Sciences*, *1339*(1), 125–137.

Eitam, B., Yeshurun, Y., & Hassan, K. (2013). Blinded by irrelevance: Pure irrelevance induced "blindness". *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 611.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67.

Jiang, Y. V., Shupe, J. M., Swallow, K. M., & Tan, D. H. (2016). Memory for recently accessed visual attributes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1331–1337.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–16.

Mack, A., & Rock, I. (1998). *Inattentional blindness.* Cambridge, MA: MIT Press.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., ... Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision, 2*(3), 191.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609–626.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 855–874.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127.

Swallow, K. M., Jiang, Y. V., & Tan, D. H. (2017). How memory is tested influences what is measured: Reply to Wyble and Chen (2017). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 1001–1003.

Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics*, *76*(7), 2136–2157.

Swan, G., Wyble, B., & Chen, H. (2017). Working memory representations persist in the face of unexpected task alterations. *Attention, Perception, & Psychophysics*, *79*(5), 1408–1414.

Theeuwes, J., Belopolsky, A., & Olivers, C. N. L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica, 132*(2), 106–114.

Wyble, B., & Chen, H. (2017). Memory consolidation of attended information is optional: Comment on Jiang et al. (2016). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 997–1000.