Published in final edited form as:

J Comp Psychol. 2013 August ; 127(3): 319–328. doi:10.1037/a0030496.

One-trial Memory and Habit Contribute Independently to Matching-to-Sample Performance in Rhesus Monkeys (*Macaca mulatta*)

Hsiao-Wei Tu and Robert R. Hampton

Emory University

Abstract

Multiple memory systems often act together to generate behavior, preventing a simple one-to-one mapping between cognitive processes and performance in specific tests. Process dissociation procedures (PDPs) have been adopted in both humans and monkeys to quantify one-trial memory and habit, with the assumption that these two processes make independent contributions to performance. Violations of this independence assumption could produce artificial dissociations. Evidence for independence has been reported in humans, but similar tests have not been conducted with monkeys until now. In a within-subjects design using a matching-to-sample task, we manipulated one-trial memory strength and habit strength simultaneously. Memory delay intervals and encoding conditions affected one-trial memory scores without affecting habit scores. In contrast, biased reinforcement selectively changed habit scores but not one-trial memory scores. This behavioral double dissociation clearly shows that one-trial memory and habit can be manipulated independently, validating PDP as a valuable tool for cross-species studies of learning and memory and reinforcing the view that one-trial memory and habits are served by distinct brain systems.

Keywords

process dissociation procedure; double dissociation; implicit memory; explicit memory; matching-to-sample

Memory is not a unitary system but rather takes different forms mediated by different brain regions (Henke, 2010; Poldrack & Packard, 2003; Sherry & Schacter, 1987; Squire & Zola-Morgan, 1991; Wixted, 2011). Some memories are formed rapidly as the result of a single specific experience, a type we call one-trial memory in this paper. By contrast, habits are acquired gradually, over many individual experiences. Neurophysiological evidence in animals and amnesic patients has demonstrated that a medial temporal lobe-dependent system supports one-trial memory, whereas a neostriatum-dependent system supports habit (Cohen & Eichenbaum, 1993; Henke, 2010; Squire, 1992; Tu, Hampton, & Murray, 2011). However, it can be difficult to get unbiased measurements of individual memory systems directly from specific cognitive tests because most test performances are the result of multiple processes (Kim & Baxter, 2001; Poldrack, et al., 2001; Poldrack & Packard, 2003).

Correspondence concerning this article should be addressed to Hsiao-Wei Tu, 109 Williams Hall (0436), Blacksburg, VA 24061. Phone: 1-540-231-9538. Fax: 1-540-231-3652. hsiaowei@vt.edu.

Author Note

Hsiao-Wei Tu and Robert R. Hampton, Department of Psychology and Yerkes National Primate Research Center, Emory University. Hsiao-Wei Tu is now at Psychology Department, Virginia Polytechnic Institute and State University.

In humans, some memories are subject to cognitive control and accessible to introspection, whereas habits are automatic and inaccessible to monitoring (e.g., Squire, Knowlton, & Musen, 1993; Squire & Zola, 1996). Thus, a commonly used method to measure these two kinds of memories is to have subjects introspect and then verbally report their private experience of memory in cognitive tests (“I remembered” vs. “I guessed”). While these reports can provide rich accounts of cognitive processing, they are also vulnerable to distortion. For example, use of cognitively controlled memory may be underreported by subjects, and the associated performance improperly interpreted as “automatic” (Richardson-Klavehn & Bjork, 1988). Subjects may also switch strategies during a task without noticing (Jacoby, 1991), making their narration of their cognitive processes unreliable.

In nonhuman animals, verbal reports are simply not available to aid in characterizing cognitive processes. Delayed matching-to-sample has been used to measure one-trial memory and object discrimination has been used to measure habits. While performance in these tests does depend disproportionately on memory and habit, an exclusive one-to-one mapping between memory processes and these tasks is unlikely (Buffalo, Stefanacci, Squire, & Zola, 1998; Hood, Postle, & Corkin, 1999). For example, if an image is used repeatedly in a session of discrimination task, memories of the previous occurrences of that image may contaminate the gradual formation of habits and prevent accurate measurement of contribution of automatic, habitual processes to performance. Similarly, previous rewarded experience with a given stimulus, or similar stimuli, may allow habits to contribute to choice in matching to sample tests.

Process dissociation procedures (PDPs) offer a valuable tool specifically designed to disentangle the involvement of two memory systems in a single task (Hay & Jacoby, 1996; Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993). By creating situations in which one-trial memory and habits are congruent and situations in which they are incongruent, the patterns of errors made by subjects are used to derive a one-trial memory score and a habit score, quantifying the contributions of these two processes. PDP designs may be especially useful in comparing human and nonhuman cognition because they do not depend on language allowing similar tests to be conducted in many species. One critical assumption of PDP is that one-trial memory and habit contribute independently to performance (Jacoby, 1991; Kelley & Jacoby, 2000). Violations of this independence assumption may result in artificial dissociation between the measures (e.g., Curran & Hintzman, 1995; Joordens & Merikle, 1993). For example, Curran & Hintzman (1995) found that increasing the presentation duration of stimuli strengthened one-trial memory as expected, but paradoxically decreased the contribution of habit at the same time. However, Jacoby et al. (1997) have argued that such “artificial dissociations” are mainly caused by violations of other conditions required for PDP, rather than by non-independence of processes.

Recently, we used PDP in a modified matching-to-sample paradigm to study the neurobiology of one-trial memory and habits in nonhuman primates (Tu, et al., 2011). This first use of PDP with monkeys took advantage of the fact that PDP does not require verbal reports, and therefore allows us to assess the behavior of animals that cannot provide verbal commentary on their private experience of memory as do humans. However, unlike the substantial support for the independence of one-trial memory and habit that exists for humans (e.g., Cowan & Stadler, 1996; Hay & Jacoby, 1996), there have been no direct behavioral tests of independence in monkeys. The present study addresses this gap, by assessing the validity of the application of PDP in nonverbal animals. This work sets the stage for further use of PDP in comparative psychology and in the development of animal models for memory research.

In Experiment 1, monkeys were trained to develop habits of varying strengths through manipulations of reinforcement history within a matching-to-sample paradigm. If habit and one-trial memory make independent contributions to choice behavior, habit should vary, reflecting multiple habit strengths, without affecting one-trial memory. To further evaluate the extent of this dissociation, we manipulated both delay length and encoding conditions in Experiment 2 in tests with the various levels of habit established in Experiment 1. If the independence assumption is valid, then the habit measures would stay constant within each level of habit strength while one-trial memory scores should vary with delay and encoding conditions.

General Method

Subjects and Apparatus

Twelve adult male rhesus monkeys (*Macaca mulatta*) were used. Monkeys were pair-housed, received a full daily food ration, and had *ad libitum* access to water. All monkeys had prior experience with automated cognitive tests using a touch-screen computer and were previously trained on clip-art visual matching-to-sample. All experiments were conducted in the monkeys' home cages. A portable testing rig consisting of a 15-inch color LCD touch-sensitive screen (Elo TouchSystems, Menlo Park, CA) running at a resolution of 1024×768 pixels and two automatic food dispensers (Med Associates, Inc., St. Albans, VT) that delivered nutritionally balanced primate pellets (Bio-Serv, Frenchtown, NJ) into food cups below the screen was attached to the front of each monkey's cage. Testing was controlled by a personal computer with custom program written in Presentation (Neurobehavioral Systems, Albany, CA). During testing, the monkeys in each pair were separated by dividers that allowed limited visual and physical contact but prevented access to one another's testing equipment. Each subject received one session per day, six days a week.

Visual Matching-to-Sample

A total of 576 color clip-art images, 160 pixels high $\times$ 200 pixels wide, were used. The images were randomly grouped into 144 "image quads," each of which contained four images that were always presented together at test. A trial began when a green square (100×100 pixels) appeared centered near the bottom of the screen. Monkeys had to touch it twice (FR2) to enter the study phase, in which the green square disappeared and one of the four images in a given image quad appeared as the sample at the center of the screen. When the subject touched the sample (FR2), the study phase ended and the whole screen turned black for a delay period. After the programmed memory delay, the test phase started with the appearance of all four images in the appropriate image quad, randomly assigned to the four corners of the screen. Touching (FR2) the image seen during the study phase of that trial was rewarded one pellet together with an "excellent!" sound. Incorrect responses resulted in a "d'oh!" sound and a 5-second timeout during which the screen stayed black. When the correction procedure was in effect, the same trial was repeated again after the timeout so that the monkey had a second chance to learn by trial and error. In the correction trials responses were reinforced as usual. If an incorrect image was chosen a second time, the same trial was presented for a third time, but during the test phase only the correct image appeared by itself in the corner where it had appeared before. Touching (FR2) this image resulted in reward as on correct trials. This ensured that these trials ended with a reinforced selection of the correct image. Consecutive trials are separated by an inter-trial interval of 5 seconds (Figure 1A).

Experiment 1: Manipulations of Habit Strength

In Experiment 1, monkeys gradually acquired “habits,” or tendencies to select one particular image in each image quad. The habits developed because some images in some quads were designated to be the sample and thus rewarded more often than the other 3 images. Habit strength was manipulated across quads by varying the extent to which selection of the sample was biased. Higher biases should be expected to produce stronger habits. However, if habit and memory are independent, variation in habit strength should not affect the strength of one-trial memory.

Method

Accuracy titration—The purpose of this stage was to adjust the visual matching performance of each monkey to between 50% and 75% correct to ensure a sufficiency of errors for the PDP calculations and to eliminate the possibility that results would be affected by ceiling or floor effects. All 144 image quads were used once only in a random order in each daily testing session and were re-used day after day. On each trial, the sample image was randomly selected from a given image quad without bias such that each of the 4 images in that quad appeared as the sample equally often across days. Monkeys were rewarded for choosing the sample image at test. Each session included 96 trials with a 3-second delay interval, and 48 probe trials with a longer memory delay (Figure 1A). The delay used for the probe trials started at 1 minute. If a monkey performed better than 75% in a block of 3 sessions, the probe delay was doubled. If performance on probe trials was lower than 50% for two consecutive blocks, we decreased the probe delay by a quarter. Monkeys progressed to next stage of testing when the average performance on long delay probe trials fell between 50% and 75% for two consecutive blocks of three sessions. The correction procedure was applied to regular trials but not probe trials.

In order to decrease performance into the criterion range without having to use extremely long delay intervals, monkeys that performed better than 75% with the 2-minute delay were tested using “clusters.” Cluster trials started with the green square like other trials. However, after touching (FR2) the sample, instead of entering the delay, another sample image appeared. The monkey had to touch (FR2) each sample image presented in that cluster in turn in order to start the delay interval. After the delay, multiple tests were presented one by one in an order corresponding to the order of sample presentations (Figure 1B). Choosing the sample image at test resulted in an “excellent!” sound and a food reward. Incorrect responses were followed by a “d’oh!” sound but no timeout interval. Omitting the time-out allowed us to keep the delay interval between the offset of each sample image and the onset of the corresponding test images constant for each test within a cluster, even when monkeys made mistakes. Clusters were separated by a 5 second inter-trial interval. Performance in each session was evaluated and accuracy was titrated by adjusting delay length and cluster size alternately, starting with a cluster of 2 and delay of 1 minute. If accuracy exceeded 75%, either the delay was increased by 1 minute or the cluster size was increased. Clusters were increased by 1 or 2 images at a time, depending on whether the resulting number of images in a cluster was a factor of 144, the total number of image quads in a session. We assumed that the larger the cluster size, the more difficult it would be for the monkeys to remember all the sample images seen during the study phase. There were no regular trials and every trial used the same long delay. The correction procedure was not in effect. The different manipulations using probes and clusters throughout the study are summarized in Table 1. Monkeys proceeded to the next stage when their performance was between 50% and 75% for two consecutive sessions.

Habit formation—In this stage, monkeys developed habits of different strengths through biased matching-to-sample. Rather than randomly selecting a sample image from an image quad without bias each trial, the probability of a designated “high-frequency image” being the sample was varied across quads, creating four bias levels: 100%, 75%, 50%, and 25%. Each bias level consisted of 36 image quads, resulting in 36 “habits.” In the *100% bias level*, the high-frequency image was always the sample. With these quads, monkeys could achieve 100% accuracy by always choosing the high-frequency image at test, even though they might in fact not remember that image as the sample. In the *75% and 50% bias level*, the high-frequency image served as the sample with a probability of 0.75 and 0.50, respectively, and the other three images in each image quad served as the sample with equal probability (0.083 and 0.167, respectively). Monkeys could therefore be correct 75% and 50% of the time if they always chose the high-frequency image from these quads at test. They could achieve 100% accuracy only if they remembered the sample image seen during study phase on each trial. The *25% bias level* was the same as initial matching training. All four images in an image quad served as the sample equally often (Figure 2). Selection of images was done by drawing without replacement from pools constructed according to the specified probabilities ensuring that across sessions each image occurred exactly the intended number of times. All 36 image quads from each bias level were intermixed in each session. The delay was 3 seconds for all 144 trials in a session. There were 12 sessions in a block. The correction procedure was in effect such that all trials ended with a rewarded response, ensuring that habits would be formed as intended.

Habit strength assessment—To monitor the development of habits, empty trials were implemented in a session in two different ways for monkeys tested with probes and clusters, respectively (Table 1). On these trials, no sample was presented, thus leaving no memory trace at test. Subjects could only “guess” when confronted with the test images. Given that no one-trial memory is available to control choice on empty trials, we interpreted any bias toward selecting particular images as reflecting habits.

Monkeys tested with probes: Three habit assessment sessions were conducted after each block of 12 habit formation sessions. In each of these habit assessment sessions, 12 image quads from each bias level were selected to serve as empty probe trials. On these probe trials, the green square signaling the beginning of a trial was presented and followed immediately after FR2 by the titrated long delay, after which all four images in the quad were presented in the test phase. Monkeys were rewarded no matter which image was selected. The remaining 24 trials in each bias level were presented exactly as in habit formation sessions. The correction procedure was only used with these trials, not with probe trials. Monkeys proceeded to the next stage when the performance on the 100% biased empty probe trials was higher than 50% for two consecutive blocks.

Monkeys tested with clusters: All 144 trials were empty trials and grouped into clusters in a single habit assessment session. Trials began with a green square followed by the delay after touching. No sample images were presented in any trial. After the delay, a cluster of tests appeared one at a time. Cluster size and delay were those determined through the titration procedure described above. Monkeys were rewarded for choosing the high-frequency image if the image quad was 100%, 75%, and 50% biased, but all choices were rewarded if the image quad was 25% biased. These contingencies were different from the indiscriminate reinforcement on probe trials because here the entire session contained only empty trials. If every trial was reinforced irrespective of choice, monkeys tested with clusters might learn that they could receive a food reward in the 100%, 75%, and 50% conditions no matter which image was chosen. The reason that all choices were rewarded only in the 25% bias condition was because we did not want to train the monkeys that any

particular image was correct on empty trials. They should respond randomly in the 25% bias condition. The correction procedure was not used. Monkeys were moved to next stage when performance on the 100% biased empty probe trials was higher than 50% for two consecutive blocks.

Main experiment—In this stage monkeys were tested with two types of trials that were used to generate PDP scores. On congruent trials, the high-frequency image appeared as the sample so that one-trial memory and habit corresponded to the same response. On incongruent trials, an image other than the high-frequency image was the sample such that one-trial memory and habit corresponded to different responses. These two types of trials were implemented in a session in two ways to accommodate monkeys tested with probes and clusters (Table 1).

Monkeys tested with probes: There were 6 congruent probe trials and 6 incongruent probe trials from each bias level in a session. Probe trials had long titrated delays to avoid ceiling and floor effects, while all the other trials had the three second delay as usual. The correction procedure was only used with normal trials, not with probe trials. One block consisted of 6 sessions. Each monkey ran 2 blocks, yielding 72 congruent and 72 incongruent probe trials for each bias level for analysis.

Monkeys tested with clusters: There were 18 *congruent trials* and 18 *incongruent trials* in each bias level in a session. Bias levels and trial types were intermixed in one cluster, making the trial type unpredictable. No correction procedure was used. One block consisted of 2 sessions. Each monkey was tested for 2 blocks, yielding 72 congruent and 72 incongruent probe trials on each bias level for analysis.

Data analysis—The probability that monkeys selected the high-frequency images, $p(hf)$, on congruent trials and incongruent trials respectively, was directly recorded. Note that $p(hf)$ is not equivalent to accuracy; sometimes it is the correct response and sometimes it is an error. According to the logic of PDP (Jacoby, 1991; Jacoby, et al., 1993), on congruent trials in which the sample was the high-frequency image, monkeys could give a correct response at test either by remembering the sample image from the study phase or by choosing based on habit when they failed to remember. Therefore the probability of correctly choosing the high-frequency image at test in the congruent condition can be described as

$$p(hf)_{congruent} = \text{one-trial memory} + \text{habit} (1 - \text{one-trial memory})$$

On incongruent trials in which the sample was *not* the high-frequency image, choice of the high-frequency image at test is an error and should only occur when monkeys forgot which image was the sample and chose the image by habit. Thus the probability of incorrectly choosing the high-frequency image at test in the incongruent condition can be described as

$$p(hf)_{incongruent} = \text{habit} (1 - \text{one-trial memory})$$

By subtracting these two equations, the common term “habit (1 – one-trial memory)” was cancelled and a one-trial memory score was obtained.

$$\text{One-trial memory} = p(hf)_{congruent} - p(hf)_{incongruent}$$

This one-trial memory score was then entered into the formula for $p(hf)_{incongruent}$ to obtain a habit score.

$$\text{Habit} = p(hf)_{incongruent} / 1 - (p(hf)_{congruent} - p(hf)_{incongruent})$$

The derived PDP scores were not direct measures of accuracy but estimates of the contribution of each memory system in the same task. They were the main dependent variables in each experiment and were arcsine transformed and analyzed by repeated measures ANOVAs.

Results and Discussion

Monkeys tested with probes—Six of the 12 monkeys continued to use probes with delays of one minute ($n = 2$) or two minutes ($n = 4$) for the duration of the experiment. The final titrated performance on probes was $64.12 \pm 6.9\%$ ($n = 6$). The small standard deviation indicates that the strength of one-trial memory was approximately the same for all 6 monkeys after titration.

By the end of the habit formation stage, high-frequency images were chosen on empty probe trials significantly more frequently than expected by chance (.25) at each level of habit except the 25% bias condition, which was in fact no bias: 100% mean = .64, $t_5 = 7.40$, $p = .001$, $d = 3.02$; 75% mean = .47, $t_5 = 5.38$, $p = .003$, $d = 2.20$; 50% mean = .40, $t_5 = 3.98$, $p = .011$, $d = 1.63$; 25% mean = .26, $t_5 = .50$, $p = .64$, $d = .20$. Habits were therefore successfully created. However, the probability of choosing high-frequency images on empty probe trials was significantly lower than the corresponding bias level: 100% ($t_5 = 11.59$, $p < .001$, $d = 4.73$) and 75% ($t_5 = 6.88$, $p < .001$, $d = 2.81$) and came close to matching the corresponding bias level only at 50% ($t_5 = 2.55$, $p = .05$, $d = 1.04$). Because our performance criterion only required that monkeys selected high-frequency images on empty probe trials more than 50% of the time at 100% bias level, it is likely that more sessions of habit training would have resulted in stronger habits closer to probability matching.

PDP measures were derived from congruent and incongruent probes (Table 2). Habit scores, but not one-trial memory scores, varied significantly as a function of bias condition (Figure 3; repeated measures two-way ANOVA, Process $F_{1,5} = .00$, $p = .98$, $\eta_p^2 = .00$; Bias level $F_{3,15} = 9.88$, $p = .001$, $\eta_p^2 = .66$; Process $\times$ Bias level $F_{3,15} = 10.57$, $p = .001$, $\eta_p^2 = .68$). Tests conducted on habit scores and one-trial memory scores separately showed that habit scores differed significantly (repeated measures ANOVA, $F_{3,15} = 13.18$, $p < .001$, $\eta_p^2 = .73$) but one-trial memory scores did not (repeated measures ANOVA, $F_{3,15} = .24$, $p = .87$, $\eta_p^2 = .05$). Post-hoc Least Significant Difference tests showed that 100% biased habit scores were significantly higher than 75% biased habit scores ($p = .03$), which in turn were significantly higher than 50% biased habit scores ($p = .005$), but 50% biased habit scores were not significantly higher than 25% biased habit scores ($p = .32$). Overall, the results show that habit strength can be manipulated without changing one-trial memory strength, indicated that the two processes are independent. Furthermore, habit measures in PDP were significantly correlated with the probability of selecting the high-frequency image on empty probe trials ($r = .80$, $p < .001$), providing converging evidence that the PDP habit score measures derived from congruent and incongruent trials are valid measures of habit.

Monkeys tested with clusters—Six monkeys reached the criterion of 75% correct on 2 minute probe trials and were therefore tested using clusters. Different combinations of cluster size and delay duration were used depending on individual performance ($n = 1$: cluster size of 2 and delay of 1 min; $n = 2$: cluster size of 3 and delay of 1 min; $n = 2$: cluster size of 4 and delay of 2 min; $n = 1$: cluster size of 8 and delay of 3 min). The final titrated performance was $70.14 \pm 2.28\%$ ($n = 6$). By the end of the habit formation stage, the probability of choosing high-frequency images was significantly higher than chance (.25) at each level of habit except 25% bias condition, which was in fact no bias: 100% mean = .72, $t_5 = 12.59$, $p < .001$, $d = 5.14$; 75% mean = .57, $t_5 = 8.40$, $p < .001$, $d = 3.43$; 50% mean = .48, $t_5 = 6.32$, $p = .001$, $d = 2.58$; 25% mean = .24, $t_5 = .51$, $p = .63$, $d = .21$. The data were

similar and did not differ significantly from those resulting from probes (one-trial memory scores: 100% biased, $t_{10} = .24$, $p = .81$, $d = .14$; 75% biased, $t_{10} = .69$, $p = .51$, $d = .40$; 50% biased, $t_{10} = .46$, $p = .66$, $d = .26$; 25% biased, $t_{10} = .48$, $p = .64$, $d = .28$; habit scores: 100% biased, $t_{10} = .40$, $p = .70$, $d = .23$; 75% biased, $t_{10} = 1.52$, $p = .16$, $d = .88$; 50% biased, $t_{10} = 1.69$, $p = .12$, $d = .98$; 25% biased, $t_{10} = .97$, $p = .36$, $d = .56$), showing that the cluster method is a good alternative and can be used to efficiently study long-duration memory while avoiding extremely long testing sessions.

The performances on congruent and incongruent trials were sorted across clusters and used to calculate PDP measures (Table 2). Habit measures at the four bias levels were significantly different than each other (repeated measures ANOVA, $F_{3, 15} = 44.75$, $p < .001$, $\eta_p^2 = .90$; post-hoc Least Significant Difference tests, $p = .03$ between 100% and 75%; $p = .02$ between 75% and 50%; $p = .004$ between 50% and 25%), but one-trial measures showed no significant difference (repeated measures ANOVA, $F_{3, 15} = .16$, $p = .92$, $\eta_p^2 = .03$). This is also indicated in the significant interaction between one-trial memory scores and habit scores across 4 bias levels (Figure 3; repeated measures two-way ANOVA, Process $F_{1, 5} = 6.35$, $p = .05$, $\eta_p^2 = .56$; Bias level $F_{3, 15} = 39.40$, $p < .001$, $\eta_p^2 = .89$; Process $\times$ Bias level $F_{3, 15} = 24.63$, $p < .001$, $\eta_p^2 = .83$). Again, the data showed that habits were manipulated without changing one-trial memory. Lastly, habit scores were significantly correlated with the probability of selecting the high-frequency image on empty trials reflecting the biased reinforcement history during training ($r = .89$, $p < .001$).

Experiment 2: Manipulations of Both Habit and One-Trial Memory

A behavioral single dissociation between one-trial memory and habit was demonstrated in Experiment 1. Different levels of habit had no effect on one-trial memory. However, it is possible that the independence of habit and one-trial memory is unidirectional. Although habit can be manipulated while one-trial memory stays the same, the reverse may not obtain. To further evaluate whether the two processes are independent, we tested for a double dissociation in Experiment 2.

In addition to the four bias levels developed in Experiment 1, memory delay was also manipulated, creating different strengths of one-trial memory within each habit level. If habits and one-trial memory are independent, habit scores should vary based on bias levels and one-trial memory scores vary according to the length of delays, but they should not affect each other.

Method—The same 12 monkeys used in Experiment 1 were used again, six tested with probes and six tested with clusters. In order for each monkey to finish one session per day in a reasonable time, only two bias levels and two delays were tested with probes, but all four bias levels and three delay lengths were tested with clusters (Table 1; see below for details). PDP analysis and the statistics were the same as those adopted in Experiment 1.

Because the performance on unbiased trials had been titrated in Experiment 1, we did not go through the whole titration process again. Instead, the overall performance on probes of 25% biased quads was evaluated. Those trials were in fact unbiased and were conducted using the titrated delay. If the performance of those trials was higher than 75%, the titrated delay was increased by a factor of 1.5. If the performance was lower than 50%, the titrated delay was decreased by a factor of 0.75. There was no habit formation stage, and no sessions for assessing habit strength, because the same image quads, and therefore the same habits, from Experiment 1 were used again.

Monkeys tested with probes: To reduce the number of conditions and thus the total duration of a session, only two bias levels and two delay intervals were used. The bias levels

were 100% and 25%, and the two delays were half the titrated delay and double the titrated delay, resulting in four conditions in a session. There were 72 trials per session, including four congruent and four incongruent probe trials in each condition and 40 regular three second delay trials. The correction procedure was in effect only for regular trials. Nine sessions constituted a block, and each monkey completed two blocks to produce 72 congruent and 72 incongruent probe trials in each condition for analysis.

Monkeys tested with clusters: All four bias levels, 100%, 75%, 50%, and 25%, were used, each of which was crossed with 3 different delays, the original titrated delay, half the titrated delay, and double the titrated delay, resulting in 12 conditions in a session. There were 6 congruent and 6 incongruent trials in each condition, making a total of 144 trials in a session. Tests of different bias levels were intermixed in a cluster, but all tests in the same cluster used the same delay. The correction procedure was not in use. Six sessions made up a block, and each monkey ran 2 blocks to produce 72 congruent and 72 incongruent trials in each condition for analysis.

Results and Discussion

Monkeys tested with probes: Two monkeys had improved and were given a longer delay as their “titrated delay” (3 min and 1.5 min, respectively); one monkey went back to use a shorter titrated delay (45 sec); the other three monkeys stayed on the same titrated delay (2 min) used in Experiment 1.

PDP measures were derived from congruent and incongruent probes (Table 3). One-trial memory scores changed significantly across different delay lengths but not across bias levels (Figure 4A; repeated measures two-way ANOVA, Bias level $F_{1,5} = .05, p = .84, \eta_p^2 = .01$; Delay $F_{1,5} = 183.08, p < .001, \eta_p^2 = .97$; Bias level $\times$ Delay $F_{1,5} = .00, p = 1.00, \eta_p^2 = .00$), whereas habit scores changed significantly across bias levels but not across different delay lengths (Figure 4B; repeated measures two-way ANOVA, Bias level $F_{1,5} = 43.73, p = .001, \eta_p^2 = .90$; Delay $F_{1,5} = .10, p = .76, \eta_p^2 = .02$; Bias level $\times$ Delay $F_{1,5} = 1.40, p = .29, \eta_p^2 = .22$). This pattern is also indicated in the significant interaction among memory processes, bias levels, and delay lengths (Figure 4; repeated measures three-way ANOVA, Bias level $F_{1,5} = 73.00, p < .001, \eta_p^2 = .94$; Delay $F_{1,5} = 61.76, p = .001, \eta_p^2 = .93$; Process $F_{1,5} = .09, p = .78, \eta_p^2 = .02$; Bias level $\times$ Delay $F_{1,5} = 1.76, p = .24, \eta_p^2 = .26$; Bias level $\times$ Process $F_{1,5} = 25.89, p = .004, \eta_p^2 = .84$; Delay $\times$ Process $F_{1,5} = 34.54, p = .002, \eta_p^2 = .87$; Bias level $\times$ Delay $\times$ Process $F_{1,5} = .64, p = .46, \eta_p^2 = .11$). These results strongly support the hypothesis that one-trial memory and habit are independent processes.

Monkeys tested with clusters: All 6 monkeys continued to use the same combinations of cluster size and delay as used in Experiment 1. Data from all clusters were pooled and then sorted by condition to derive PDP measures (Table 3). Across the three delay lengths, one-trial memory scores varied significantly but not habit scores (Figure 5A; repeated measures two-way ANOVA, Bias level $F_{3,15} = 1.49, p = .26, \eta_p^2 = .23$; Delay $F_{2,10} = 109.28, p < .001, \eta_p^2 = .96$; Bias level $\times$ Process $F_{6,30} = .78, p = .60, \eta_p^2 = .13$). Across 4 bias levels, habit scores varied significantly but not one-trial memory scores (Figure 5B; repeated measures two-way ANOVA, Bias level $F_{3,15} = 1.74, p = .17, \eta_p^2 = .23$; Delay $F_{2,10} = 109.28, p < .001, \eta_p^2 = .96$; Bias level $\times$ Process $F_{6,30} = .78, p = .60, \eta_p^2 = .13$). Moreover, the interaction among memory processes, bias levels, and delay lengths was significant (Figure 5; repeated measures three-way ANOVA, Bias level $F_{3,15} = 28.85, p < .001, \eta_p^2 = .85$; Delay $F_{2,10} = 32.73, p < .001, \eta_p^2 = .87$; Process $F_{1,5} = .00, p = .98, \eta_p^2 = .00$; Bias level $\times$ Delay $F_{6,30} = 1.01, p = .44, \eta_p^2 = .17$; Bias level $\times$ Process $F_{3,15} = 9.19, p = .001, \eta_p^2 = .65$; Delay $\times$ Process $F_{2,10} = 16.57, p = .001, \eta_p^2 = .77$; Bias level $\times$ Delay $\times$ Process $F_{6,30} = 1.18, p = .35, \eta_p^2 = .19$). These data provided converging evidence that one-trial

memory and habit are independent processes. Our results also show that the clusters method yields results similar to those acquired from probes, while allowing us to test more conditions within a single session.

Discussion

Using PDP, we observed a behavioral double dissociation between one-trial memory and habit in monkeys, demonstrating reliable methods for measuring the contributions of these two processes. In animal studies, delayed matching-to-sample has been used to measure one-trial memory (Mishkin, Malamut, & Bachevalier, 1984), based on the premise that longer delays reduce accuracy because memory traces decay with the passage of time (Grant & Roberts, 1973). On the other hand, concurrent discrimination tasks, in which subjects gradually learn to choose one particular object through repetitive reinforcement, have been used to measure habits (Mishkin, et al., 1984), but different parameters in these experiments may cause different degrees of contamination from other memory processes. For example, frequent repetition of stimuli during habit acquisition (e.g., 8 pairs of objects, each repeated 5 times in a session per day; Suzuki, Zola-Morgan, Squire, & Amaral, 1993) may permit one-trial memories of specific trial outcomes to persist long enough to contribute to performance on subsequent trials, thereby contaminating the measure of habit with one-trial memory. Although this effect could be reduced by prolonging the time interval between image repetition (e.g., reuse the same images every 24 hours; Malamut, Saunders, & Mishkin, 1984), in many cases it would still be difficult to gauge whether more than one memory process is involved. PDP is especially important because one-trial memory and habit are derived from performance in the same context, providing ideal control for a variety of factors that might vary between distinct tasks used to measure the processes separately.

PDP has been used frequently to dissociate two concurrently acting processes, one under cognitive control and one not (McBride, 2007), with the assumption that the two processes being measured are independent (Jacoby, 1991). Although it is debated whether this independence assumption holds true in every situation (e.g., Curran & Hintzman, 1995; Joordens & Merikle, 1993), the present study demonstrated independence in two different ways. First, we evaluated the correspondence between measures from PDP and those from traditional tests that are as uncontaminated as possible. The empty trials included in our experiments resembled object discrimination tasks typically used to measure habits in nonhuman animals and served as a stand-alone, direct measure of habit strength. Performance on these trials right before PDP manipulations was highly correlated with the habit scores obtained later, which showed strong evidence that habit is independent of one-trial memory. Second, we found a double dissociation in Experiment 2 when one-trial memory and habit were manipulated simultaneously, providing direct support that one-trial memory and habit act independently. In a human study, manipulations of study duration were reflected in one-trial memory scores but not habit scores. In a different experiment, habits were established through biased training and the strength of the habits was reflected in habit scores but not one-trial memory scores (Hay & Jacoby, 1996). However, it is possible that these single dissociations may be simply due to the fact that one process is more easily affected by certain factors than the other, not that they are completely independent. Our data clearly showed that in the same task, higher bias levels resulted in higher habit scores and vice versa but one-trial memory scores always remained constant, whereas longer memory delays impaired one-trial memory scores but spared habit scores.

Process dissociation procedure offers a behavioral method to quantitatively separate the contributions of one-trial memory and habit in a single cognitive task (Jacoby, 1991; Jacoby, et al., 1993). Typically in human studies, specific instructions are given for the subjects to use their memory (e.g., “try to complete the word stems with a previously studied word”) in

an “inclusion test” or to withhold what was remembered (e.g., “try *not* to complete the word stems with any previously studied word”) in an “exclusion test” (Jacoby, 1991). We obviously cannot instruct the monkeys in the same way. It is also generally more difficult to train an animal to withhold than to express a behavior, which may make the animal more willing to perform in inclusion conditions than in exclusion conditions. Therefore, we adopted another method, in which the two situations where one-trial memory and habit either cooperate or compete were created by manipulating stimulus congruency. The monkeys were trained to always choose the image seen during the study phase in a given trial, the same instruction in regular matching-to-sample tasks. This eliminates the problems associated with using different instructions to create experimental conditions in nonhuman animals and also ensures that motivation is constant across tests.

The current experiments included several improvements over the first, and to our knowledge the only existing use of PDP in monkeys (Tu, et al., 2011). First, one-trial memory was equated across monkeys using delay titration before PDP manipulation. Like humans, the ability to remember an image varies among monkeys. By titrating the performance on unbiased matching-to-sample tasks, each monkey started the main experiment with an individually tailored memory interval and cluster size so that all subjects had similar one-trial memory scores. This step also helped to avoid ceiling effects that would violate the test conditions necessary for PDP (Jacoby, et al., 1997). Second, the inclusion of empty trials during habit training offered a method to directly monitor the development of habits. Lastly, “clusters” were used in the present research. Practically, clusters provide a more efficient way to include more variations in complicated experiments like those reported here. We obtained similar results with clusters and with probes, which encourages use of clusters in future work. Theoretically, clusters are more closely analogous to the methods typically used in human research, in which subjects are required to study a list of items and then tested with a list of “exclusion” or “inclusion” tasks (Jacoby, 1991; Jacoby, Woloshyn, & Kelley, 1989). Therefore, unlike trial-by-trial tests usually seen in animal studies, clusters offer an opportunity to make more direct comparisons between human and nonhuman memory systems.

Behavioral dissociation between controlled and automatic memory is important since multiple memory systems act simultaneously under most natural situations. It is also critical in the development of animal models for understanding the neurobiological mechanisms supporting different memory systems in humans because behavioral responses are the ultimate outcome available for examination. Here we showed that one-trial memory and habit are two independent processes in nonverbal monkeys, which addressed some criticisms on the independence assumption of PDP. This work further confirmed that PDP is a valuable behavioral paradigm appropriate for establishing correspondences between memory systems in humans and animals and set up the foundation for future investigations on the interactions among different cognitive processes in nonverbal animals.

Acknowledgments

This work was supported by National Institute of Mental Health Grant No. 1R01MH082819, National Science Foundation Grant No. 0745573. This project was funded by the National Center for Research Resources P51RR165 and is currently supported by the Office of Research Infrastructure Programs/OD P51OD11132. We thank Emily K. Brown and Steven L. Sherrin for assistance with testing animals and animal care.

References

- Buffalo EA, Stefanacci L, Squire LR, Zola SM. A reexamination of the concurrent discrimination learning task: The importance of anterior inferotemporal cortex, area TE. *Behavioral Neuroscience*. 1998; 112:3–14. [PubMed: 9517811]

- Cohen, NJ.; Eichenbaum, HE. Memory, amnesia, and the hippocampal system. MIT Press; Cambridge, MA: 1993.
- Cowan T, Stadler MA. Estimating unconscious processes: Implications of a general class of models. *Journal of Experimental Psychology: General*. 1996; 125:195–200. [PubMed: 8683194]
- Curran T, Hintzman DL. Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21:531–547.
- Grant DS, Roberts WA. Trace interaction in pigeon short-term memory. *Journal of Experimental Psychology*. 1973; 101:21–29.
- Hay JF, Jacoby LL. Separating habit and recollection: Memory slips, process dissociations, and probability matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1996; 22:1323–1335.
- Henke K. A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*. 2010; 11:523–532.
- Hood KL, Postle BR, Corkin S. An evaluation of the concurrent discrimination task as a measure of habit learning: performance of amnesic subjects. *Neuropsychologia*. 1999; 37:1375–1386. [PubMed: 10606012]
- Jacoby LL. A process dissociation framework: Separating automatic from intentional use of memory. *Journal of Memory and Language*. 1991; 30:513–541.
- Jacoby LL, Begg IM, Toth JP. In defense of functional independence: Violations of assumptions underlying the process-dissociation procedure? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1997; 23:484–495.
- Jacoby LL, Toth JP, Yonelinas AP. Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*. 1993; 122:139–154.
- Jacoby LL, Woloshyn V, Kelley C. Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*. 1989; 118:115–125.
- Joordens S, Merikle PM. Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General*. 1993; 122:462–467.
- Kelley, CM.; Jacoby, LL. Recollection and familiarity. In: Tulving, E.; Craik, FIM., editors. *The Oxford handbook of memory*; Oxford UP; New York: 2000. p. 215–228.
- Kim JJ, Baxter MG. Multiple brain-memory systems: the whole does not equal the sum of its parts. *Trends in Neurosciences*. 2001; 24:324–330. [PubMed: 11356503]
- Malamut B, Saunders R, Mishkin M. Monkeys with combined amygdalo-hippocampal lesions succeed in object discrimination learning despite 24-hour intertrial intervals. *Behavioral Neuroscience*. 1984; 98:759–769. [PubMed: 6487412]
- McBride DM. Methods for measuring conscious and automatic memory. *Journal of Consciousness Studies*. 2007; 14:198–215.
- Mishkin, M.; Malamut, B.; Bachevalier, J. Memories and habits: Two neural systems. In: Lynch, G.; McGaugh, JL.; Weinberger, NM., editors. *Neurobiology of Learning and Memory*. Guilford Press; New York: 1984.
- Poldrack RA, Clark J, Paré-Blagoev EJ, Shohamy D, Moyano JC, Myers C, et al. Interactive memory systems in the human brain. *Nature*. 2001; 414:546–550. [PubMed: 11734855]
- Poldrack RA, Packard MG. Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*. 2003; 41:245–251. [PubMed: 12457750]
- Richardson-Klavehn A, Bjork RA. Measures of memory. *Annual Review of Psychology*. 1988; 39:475–543.
- Sherry DF, Schacter DL. The evolution of multiple memory systems. *Psychological Review*. 1987; 94:439–454.
- Squire LR. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review*. 1992; 99:195–231. [PubMed: 1594723]
- Squire LR, Knowlton B, Musen G. The structure and organization of memory. *Annual Review of Psychology*. 1993; 44:453–495.

- Squire LR, Zola-Morgan S, Zola-Morgan S, Squire LR. The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience*. 2011; 34:259–288.
- Squire LR, Zola SM. Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93:13515–13522. [PubMed: 8942965]
- Suzuki WA, Zola-Morgan S, Squire LR, Amaral DG. Lesions of the perirhinal and parahippocampal cortices in the monkey produce long-lasting memory impairment in the visual and tactual modalities. *Journal of Neuroscience*. 1993; 13:2430–2451. [PubMed: 8501516]
- Tu H-W, Hampton RR, Murray EA. Perirhinal cortex removal dissociates two memory systems in matching-to-sample performance in rhesus monkeys. *Journal of Neuroscience*. 2011; 31:16336–16343. [PubMed: 22072685]

Figure 1.

(A) An example of a visual matching-to-sample trial. On probe trials, the delay was adjusted to create different levels of one-trial memory strength. However, it is not an efficient method when the delay is much longer than a few seconds in order to decrease memories of the sample in monkeys. Moreover, probe trials are only a small proportion of a session to prevent learning. (B) An example of a cluster with a size of 3 trials. Three sample images were presented one after another by touching (FR2). The delay was calculated so that the time elapsed between the offset of each sample and the onset of the corresponding test was carefully controlled. Because the whole session was divided into multiple clusters and every trial was functionally a “probe,” performance could be titrated within fewer sessions.

Figure 2.

The sample image in each trial was chosen according to the bias level that image quad belonged to. There were 36 image quads in each bias level, but only one per bias level was shown here as an example. Each image quad was used once per day. The high-frequency image served as the sample for 100%, 75%, 50%, and 25% of the time, respectively. The other 3 non-high-frequency distractors had equal chance to be the sample.

Figure 3. PDP measures of one-trial memory (A) and habit (B) in matching-to-sample tasks showed a single dissociation of these two processes both by probes (black bars) and by clusters (gray bars) in Experiment 1. Changing bias levels independently manipulated habit scores and kept one-trial memory scores constant. Error bars indicate 95% confidence interval.

Figure 4.

PDP measures of one-trial memory (A) and habit (B) in matching-to-sample tasks showed a double dissociation of these two processes by probes in Experiment 2. One-trial memory scores varied according to delay lengths but stayed constant across bias levels. Habit scores varied according to bias levels but stayed constant across delay lengths. Delay lengths are shown in the figure legend. Error bars indicate 95% confidence interval.

Figure 5. PDP measures of one-trial memory (A) and habit (B) in matching-to-sample tasks showed a double dissociation of these processes by clusters in Experiment 2. One-trial memory scores varied according to delay lengths but stayed constant across bias levels. Habit scores varied according to bias levels but stayed constant across delay lengths. Delay lengths are shown in the figure legend. Error bars indicate 95% confidence interval.

Table 1

A summary of the different steps for monkeys tested with probes and clusters in the present study. Note that the procedure for habit formation was identical without using probes or clusters.

Using PROBES	Using CLUSTERS
<u>Performance titration</u>	
• 48 probe trials with a delay length <1 min	• 144 trials divided into clusters
<u>Habit formation</u>	
<i>4 bias levels</i>	<i>4 bias levels</i>
• 3-sec delay for all 36 trials per bias level	• 3-sec delay for all 36 trials per bias level
<u>Habit strength assessment</u>	
• 12 empty probe trials per bias level	• 36 empty trials per bias level
<u>Experiment 1</u>	
<i>4 bias levels</i>	<i>4 bias levels</i>
• 6 congruent probe trials per bias level	• 18 congruent trials per bias level
• 6 incongruent probe trials per bias level	• 18 incongruent trials per bias level
<u>Experiment 2</u>	
<i>2 bias levels × 2 delay lengths = 4 conditions</i>	<i>4 bias levels × 3 delay lengths = 12 conditions</i>
• 4 congruent probe trials per condition	• 6 congruent trials per condition
• 4 incongruent probe trials per condition	• 6 incongruent trials per condition

Table 2

Probability (mean $\pm$ 95% CI) of choosing the high-frequency image on congruent and incongruent trials as probes or clusters under various conditions in Experiment 1. The value is the same as percent correct on congruent trials but is the probability of erring by choosing the high-frequency image on incongruent trials.

Bias level	Using Probes		Using Clusters	
	Congruent	Incongruent	Congruent	Incongruent
100%	0.90 $\pm$ 0.06	0.39 $\pm$ 0.14	0.90 $\pm$ 0.02	0.41 $\pm$ 0.10
75%	0.80 $\pm$ 0.08	0.27 $\pm$ 0.08	0.84 $\pm$ 0.05	0.35 $\pm$ 0.04
50%	0.71 $\pm$ 0.06	0.21 $\pm$ 0.08	0.76 $\pm$ 0.08	0.31 $\pm$ 0.07
25%	0.66 $\pm$ 0.16	0.13 $\pm$ 0.06	0.59 $\pm$ 0.06	0.13 $\pm$ 0.02

Table 3

Probability (mean $\pm$ 95% CI) of choosing the high-frequency image on congruent and incongruent trials as probes or clusters under various conditions in Experiment 2. The value is the same as percent correct on congruent trials but is the probability of erring by choosing the high-frequency image on incongruent trials.

Bias level	Delay length	Using Probes		Using Clusters	
		Congruent	Incongruent	Congruent	Incongruent
100%	Half	0.94 $\pm$ 0.03	0.24 $\pm$ 0.09	0.89 $\pm$ 0.03	0.26 $\pm$ 0.11
	Regular	NA	NA	0.86 $\pm$ 0.03	0.26 $\pm$ 0.11
	Double	0.85 $\pm$ 0.06	0.44 $\pm$ 0.12	0.75 $\pm$ 0.05	0.34 $\pm$ 0.13
75%	Half	NA	NA	86 $\pm$ 005	24 $\pm$ 008
	Regular	NA	NA	0.83 $\pm$ 0.03	0.29 $\pm$ 0.10
	Double	NA	NA	0.77 $\pm$ 0.08	0.33 $\pm$ 0.10
50%	Half	NA	NA	82 $\pm$ 006	19 $\pm$ 006
	Regular	NA	NA	0.77 $\pm$ 0.07	0.26 $\pm$ 0.06
	Double	NA	NA	0.73 $\pm$ 0.08	0.33 $\pm$ 0.06
25%	Half	79 $\pm$ 008	0.09 $\pm$ 0.04	74 $\pm$ 010	13 $\pm$ 005
	Regular	NA	NA	0.69 $\pm$ 0.07	0.16 $\pm$ 0.05
	Double	0.60 $\pm$ 0.11	0.18 $\pm$ 0.06	0.59 $\pm$ 0.08	0.20 $\pm$ 0.06