Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement

Alicia L. DeRusso[1], David Fan[1], Jay Gupta[1], Oksana Shelest[1], Rui M. Costa[2] and Henry H. Yin[1]*

[1] Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University, Durham, NC, USA
[2] Champalimaud Neuroscience Programme, Instituto Gulbenkian de Ciência, Oeiras, Portugal

Interval schedules of reinforcement are known to generate habitual behavior, the performance of which is less sensitive to revaluation of the earned reward and to alterations in the action-outcome contingency. Here we report results from experiments using different types of interval schedules of reinforcement in mice to assess the effect of uncertainty, in the time of reward availability, on habit formation. After limited training, lever pressing under fixed interval (FI, low interval uncertainty) or random interval schedules (RI, higher interval uncertainty) was sensitive to devaluation, but with more extended training, performance of animals trained under RI schedules became more habitual, i.e. no longer sensitive to devaluation, whereas performance of those trained under FI schedules remained goal-directed. When the press-reward contingency was reversed by omitting reward after pressing but presenting reward in the absence of pressing, lever pressing in mice previously trained under FI decreased more rapidly than that of mice trained under RI schedules. Further analysis revealed that action-reward contiguity is significantly reduced in lever pressing under RI schedules, whereas action-reward correlation is similar for the different schedules. Thus the extent of goal-directedness could vary as a function of uncertainty about the time of reward availability. We hypothesize that the reduced action-reward contiguity found in behavior generated under high uncertainty is responsible for habit formation.

Keywords: interval schedule of reinforcement, basal ganglia, learning, devaluation, reward, uncertainty, degradation, omission
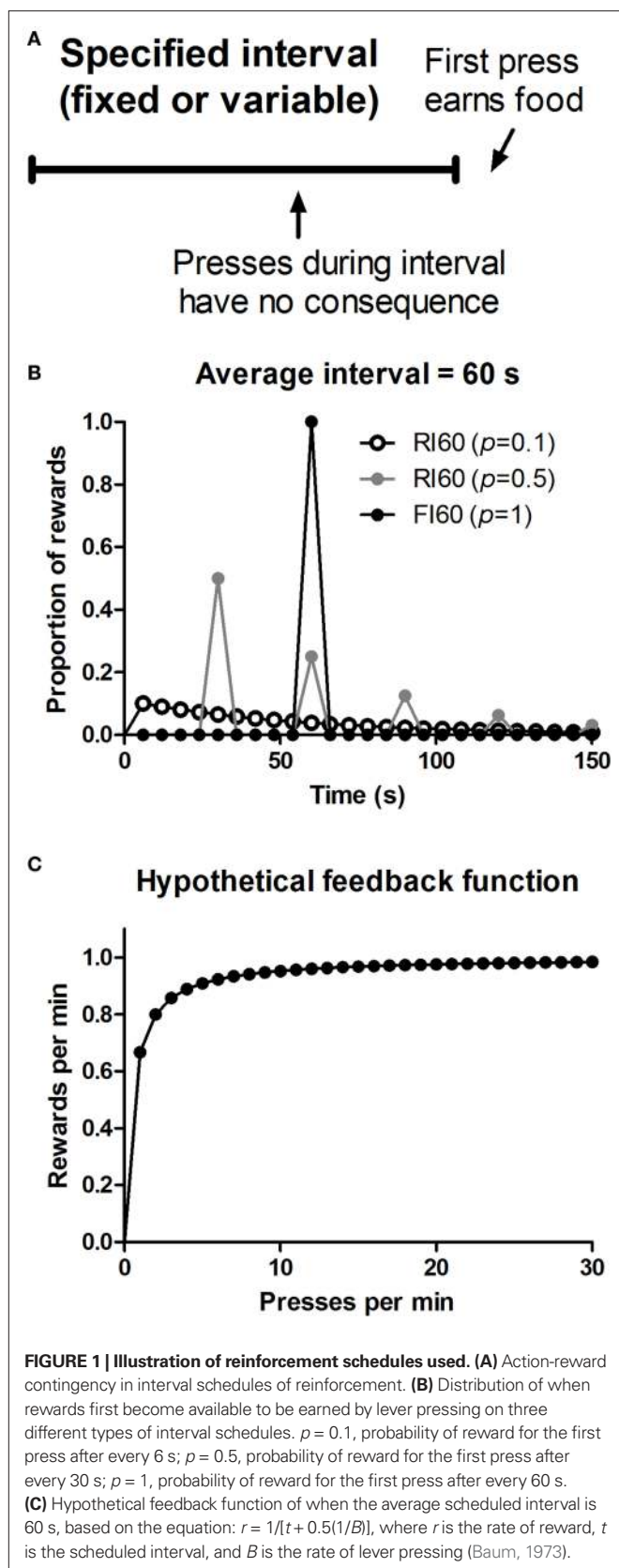
INTRODUCTION

Instrumental behavior is governed by the contingency between the action and its outcome. Under different 'schedules of reinforcement', which specify when a reward is delivered following a particular behavior, animals display distinct behavioral patterns.

In interval schedules, the first action after some specified interval earns a reward (Ferster and Skinner, 1957). Such schedules model naturally depleting resources in the environment: An action is necessary to obtain reward, but the reward is not always available – being depleted and replenished at regular intervals. Interval schedules generate predictable patterns of behavior, which have been described in detail by previous investigators (Ferster and Skinner, 1957; Catania and Reynolds, 1968).

An interesting feature of interval schedules is their capacity, under some conditions, to promote habit formation, operationally defined as behavior insensitive to updates in outcome value and action-outcome contingency (Dickinson, 1985). Studies have suggested that instrumental behavior can vary in the degree of goal-directedness. When it is explicitly goal-directed, performance reflects the current value of the outcome and the action-outcome contingency. But when it becomes more habitual, performance is independent of the current value of the goal and the instrumental contingency (Dickinson, 1985). These two modes of instrumental control can be dissociated using assays that manipulate either the outcome value or action-outcome contingency. Given one action (e.g. lever pressing) and one reward (e.g. food pellet), interval schedules are known to promote habit formation while ratio schedules do

not, even when they yield comparable rates of reward (Dickinson et al., 1983). In this respect they have been contrasted with ratio schedules, the other major class of reinforcement schedules, in which the rate of reinforcement is a monotonically increasing function of the rate of behavior. Indeed, the distinction between actions and habits was initially based on results from a direct experimental comparison between these two types of schedules (Adams and Dickinson, 1981; Adams, 1982; Colwill and Rescorla, 1986; Dickinson, 1994). In ratio schedules, the more one performs the action (e.g. presses a lever) the higher the rate of reward. But in interval schedules, the correlation between behavior and reward is more limited. Higher rates of lever pressing do not result in higher reward rates, since the reward is depleted and the feedback function quickly asymptotes (**Figure 1**). For example, under a random interval (RI) 60 schedule, the maximum reward rate is on average about one reward per minute, and cannot be increased no matter how quickly the animal presses the lever.

The ability of interval schedules to promote habit formation has been attributed to their low instrumental contingency, defined as the correlation between the reward rate and lever press rate (Dickinson, 1985, 1994). Although the reduced instrumental contingency in interval schedules is evident from their feedback functions (**Figure 1**), it is not clear whether such feedback **functions** *per se* can explain behavior (Baum, 1973). What is the time window used to detect relationships between actions and consequences? Is an animal's behavioral policy based on the correlation experienced, say, in the last hour, or in the last 10 s? These two alternatives are

FIGURE 1 | Illustration of reinforcement schedules used. (A) Action-reward contingency in interval schedules of reinforcement. **(B)** Distribution of when rewards first become available to be earned by lever pressing on three different types of interval schedules. $p = 0.1$, probability of reward for the first press after every 6 s; $p = 0.5$, probability of reward for the first press after every 30 s; $p = 1$, probability of reward for the first press after every 60 s. **(C)** Hypothetical feedback function of when the average scheduled interval is 60 s, based on the equation: $r = 1/[t + 0.5(1/B)]$, where r is the rate of reward, t is the scheduled interval, and B is the rate of lever pressing (Baum, 1973).

schedules, the interval is always the same, but in variable schedules (e.g. random interval schedules), this value can vary. Despite similar overall feedback functions, the local experienced contingency for these schedules may differ, as they generate very different behavioral patterns. In fixed interval (FI) schedules, the animal can learn to time the interval, and press more quickly towards the end of the interval, resulting in a well-known 'scalloping' pattern in the cumulative record; whereas in RI schedules the rate of lever pressing is more constant, due to the uncertainty about the time of reward availability (Ferster and Skinner, 1957).

Previous research on habit formation did not distinguish between FI and RI schedules, even though most studies used RI schedules (Yin et al., 2004). If interval uncertainty is a determinant of habit formation, then one would predict differential sensitivity to outcome devaluation and action-outcome contingency manipulations in behaviors generated by these two types of schedules. Here we compared behaviors under three types of interval schedules that differ in the uncertainty in the time of reward availability. Using outcome devaluation and instrumental contingency omission, we then compared the lever pressing under these schedules in terms of sensitivity to outcome devaluation and omission.

MATERIALS AND METHODS

ANIMALS
All experiments were conducted in accordance with the Duke University Institutional Animal Care and Use Committee guidelines. Male C57BL/6J mice purchased from the Jackson laboratory at around 6 weeks of age were used. One week after arrival, mice were placed on a food deprivation schedule to reduce their weight to ~85% of ad lib weight. They were fed 1.5–2 g of home chow each day at least 1 h after testing and training. Water was available at all times in the home cages.

INSTRUMENTAL TRAINING
Training and testing took place in six Med Associates (St. Albans, VT) operant chambers (21.6 cm L × 17.8 cm W × 12.7 cm H) housed within light-resistant and sound attenuating walls. Each chamber contained a food magazine that received Bio-Serv 14 mg pellets from a dispenser, two retractable levers on either side of the magazine, and a 3 W 24 V house light mounted on the wall opposite the levers and magazine. A computer with the Med-PC-IV program was used to control the equipment and record behavior. An infrared beam was used to record magazine entries.

INTERVAL SCHEDULES
The interval schedules used in this study were constructed based on the procedure introduced by Farmer (Farmer, 1963). The time interval is defined as the ratio between some renewing cycle T, and a constant probability of reward, p. Thus after every cycle, the reward becomes available at a specified probability. For FI schedules, p is 1, so that T equals the interval (e.g. FI 60 means after every 60 s the probability of reward availability is 100%). One can manipulate how 'random' the interval is by changing p and T, more random schedules permitting a broader distribution of reward availability (**Figure 3**). For RI 60, $p = 0.1$ schedules, $p = 0.1$ and $T = 6$, and for RI 60, $p = 0.5$ schedules, $p = 0.5$ and $T = 30$.

traditionally associated with 'molar' and 'molecular' accounts of instrumental behavior; and one way to test them is to compare fixed and variable interval schedules of reinforcement. In fixed

LEVER-PRESS TRAINING

Pre-training began with one 30-min magazine training session, during which pellets were delivered on a random time schedule on average every 60 s, in the absence of any reward. This allowed the animals to learn the location of food delivery. The next day, lever-press training began. At the beginning of each session, the house light was turned on and the lever inserted. At the end of each session, the house light turned off and the lever retracted. Initial lever-press training consisted of three consecutive days of continuous reinforcement (CRF), during which the animals received a pellet for each lever press. Sessions ended after 90 min or 30 rewards, whichever came first. After 3 CRF sessions, mice were divided into groups and trained on different interval schedules. Animals were trained 2 days on either RI 20 (pellets dispensed immediately after lever press on a random time schedule on average every 20 s) or FI 20 schedules (pellets dispensed immediately after lever press every 20 s). They were then trained for 6 days on the 60 interval schedules.

DEVALUATION TESTS

After 2 days of training on FI or RI 60 schedules, an early outcome devaluation test was conducted to determine if animals could learn the action-outcome relation under all the schedules. Animals were given the same amount of either the home 'chow' fed to them normally in their cages (valued condition/control), or the food pellet they normally earned during lever-press sessions (devalued condition). Home chow was used as a control for overall level of satiety. The mice were allowed to eat for 1 h. Immediately afterwards, they received a 5-min probe test, during which the lever was inserted but no pellet was delivered. On the second day of outcome devaluation, the same procedure was used, switching the two types of food (those that received home chow on day 1 received pellets on day 2, and vice versa).

OMISSION TEST

The animals were retrained for two daily sessions on the same schedules after the last devaluation test. They were then given the omission test, in which the instrumental contingency was reversed in an omission procedure, which tests the sensitivity of the animal to a change in the prevailing causal relationship between lever pressing and food reward. For the omission training, a pellet was delivered every 20 s without lever pressing, but each press would reset the counter and thus delay the food delivery. Animals were trained on this schedule for two consecutive days.
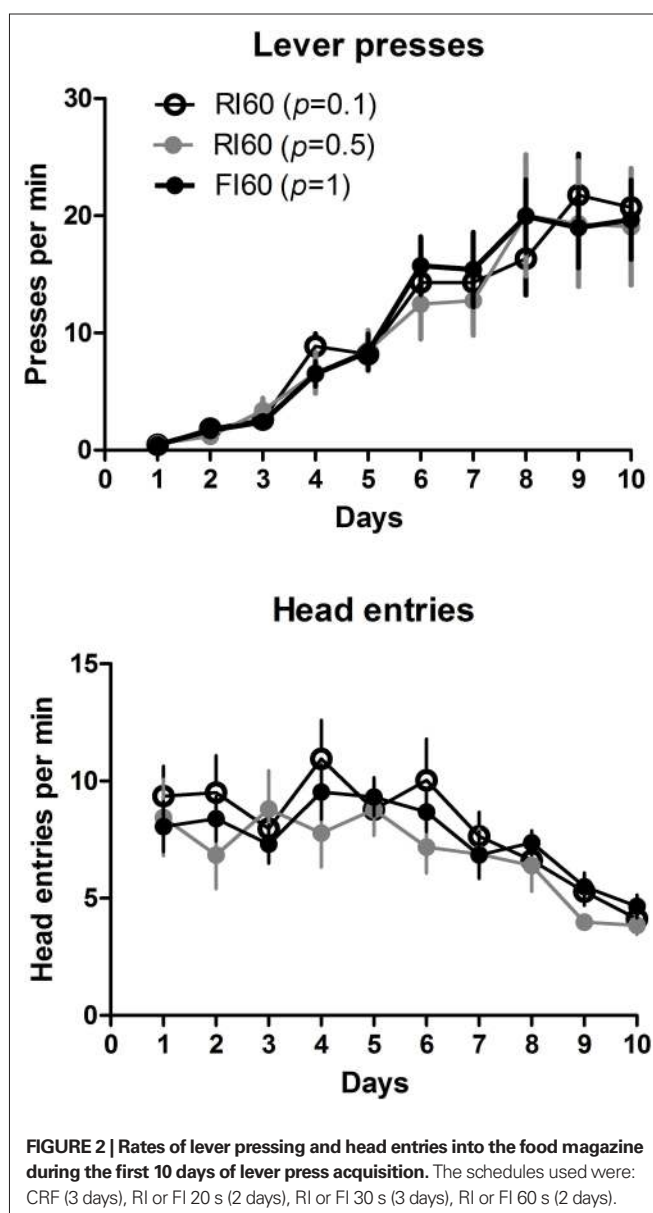
DATA ANALYSIS

Data were analyzed using Matlab, Microsoft Excel, and Prism. To calculate the local action-reward correlation, we divided the data from the last session for each animal into 60 s periods. We then divided each 60 s period into 300 bins (200 ms each). Two arrays were then created with 300 elements each, one for lever presses and the other for food pellets. Each element in a given array is the average value of press or pellet counts for a 200-ms bin. Finally the Pearson's r correlation coefficient between the press array and the pellet array was calculated. This analysis is partly based on previous work that examined action-reward correlation in humans (Tanaka et al., 2008).

To calculate the action-reward contiguity, for each lever press we measured the time between that press and the next reward.

RESULTS

INITIAL ACQUISITION

All animals learned to press the lever after three sessions of CRF training, in which each press is reinforced with a food pellet. A two-way mixed ANOVA conducted on the first 10 days of lever press acquisition (**Figure 2**), with Days and Schedule as factors, showed no interaction between these factors ($F < 1$), no effect of schedule ($F < 1$), and a main effect of Days ($F_{9,225} = 48.2, p < 0.05$), indicating that all mice, regardless of the training schedule, increased their rate of lever pressing in the first 10 days. As rate of lever pressing increased, the rate of head entries into the food magazine decreased over this period. A two-way mixed ANOVA showed no main effect of Schedule ($F < 1$), a main effect of Days ($F_{9,225} = 13.0, p < 0.05$), and no interaction between Days and Schedule ($F < 1$).

FIGURE 2 | Rates of lever pressing and head entries into the food magazine during the first 10 days of lever press acquisition. The schedules used were: CRF (3 days), RI or FI 20 s (2 days), RI or FI 30 s (3 days), RI or FI 60 s (2 days).

DEVALUATION

We conducted two outcome devaluation tests, one early in training and one after more extended training (**Figure 3**). During the early devaluation test performed after limited training (two sessions of 60-s interval schedules), rate of lever pressing in all groups decreased following specific satiety-induced devaluation relative to the control treatment (home chow). A two-way mixed ANOVA with Devaluation and Schedule as factors showed no main effect of Schedule ($F < 1$), a main effect of Devaluation ($F_{1, 25} = 19.7$, $p < 0.05$), and no interaction between these two factors. After additional training (four more sessions of 60-s interval schedules), mice that received RI training were no longer sensitive to devaluation (planned comparison $ps > 0.05$) while the FI group remained sensitive to devaluation ($p < 0.05$), showing more goal-directed behavior after extended training.

OMISSION

When the action-outcome contingency was reversed in an omission procedure, the rate of lever pressing was differentially affected in the three groups. Increasing certainty about the time of reward delivery is accompanied by increased behavioral sensitivity to the reversal of the instrumental contingency (**Figure 4**). This observation was confirmed by a one-way ANOVA: There was a main effect of Schedule ($F_{2, 25} = 10.5$, $p < 0.05$), and *post hoc* analysis showed that the rate of lever pressing is significantly higher in the RI 60 ($p = 0.1$) group compared to the FI 60 group ($p < 0.05$). At the same time, the rate of head entries to the food magazine showed the opposite pattern. There was a main effect of schedule ($F_{2, 25} = 5.04$, $p < 0.05$). *Post hoc* analysis showed that rate of head entries was significantly higher in the FI group compared to the RI group ($p < 0.05$). Thus, reduced lever pressing in the FI group is also accompanied by higher rates of head entries into the magazine. Fixed interval training, then, generated behavior significantly more sensitive to the imposition of the omission contingency.

DETAILED ANALYSIS OF LEVER PRESSING UNDER DIFFERENT INTERVAL SCHEDULES

Using Matlab, we analyzed the lever pressing under three different schedules, using data from 18 mice (6 from each group) that are run at the same time. For all analyses we used only the data from the last day of training just before the late devaluation test. **Figure 5** shows the dramatic differences in the local pattern of lever pressing under these schedules.

Mice under the three different schedules did not show significant differences in action-reward correlation. As shown in **Figure 6A**, a one-way ANOVA shows no main effect of schedule on action-reward correlation ($F < 1$). By contrast, temporal uncertainty had a significant effect on the action-reward contiguity, as shown in **Figure 6B**. A one-way ANOVA shows a main effect of schedule ($F_{2, 25} = 113$, $p < 0.05$), and *post hoc* analysis shows significant differences in all group comparisons in the time between action and reward ($ps < 0.05$).

DISCUSSION

Instrumental behavior, e.g. lever pressing for food, can become relatively insensitive to changes in outcome value or action-outcome contingency – a process known as habit formation (Dickinson,
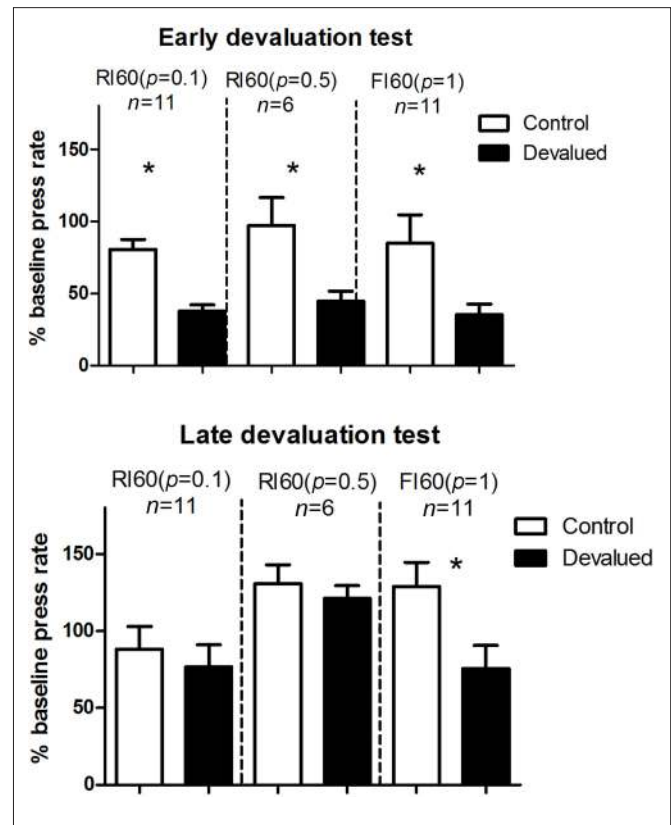

FIGURE 3 | Results from the two specific satiety outcome devaluation tests. Early devaluation, first outcome devaluation test was done after 2 days of training on the 60-s interval schedules. Late devaluation: second devaluation test was done after four additional days of training on the same 60-s schedules. For both, all mice were given a 5-min probe test conducted in extinction after specific satiety treatment.
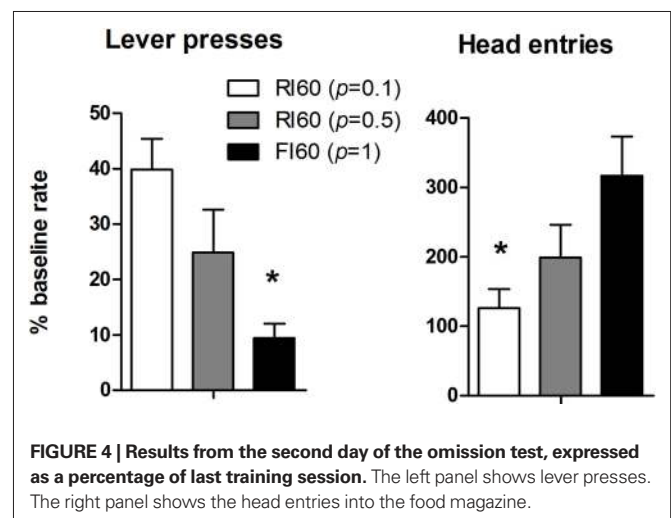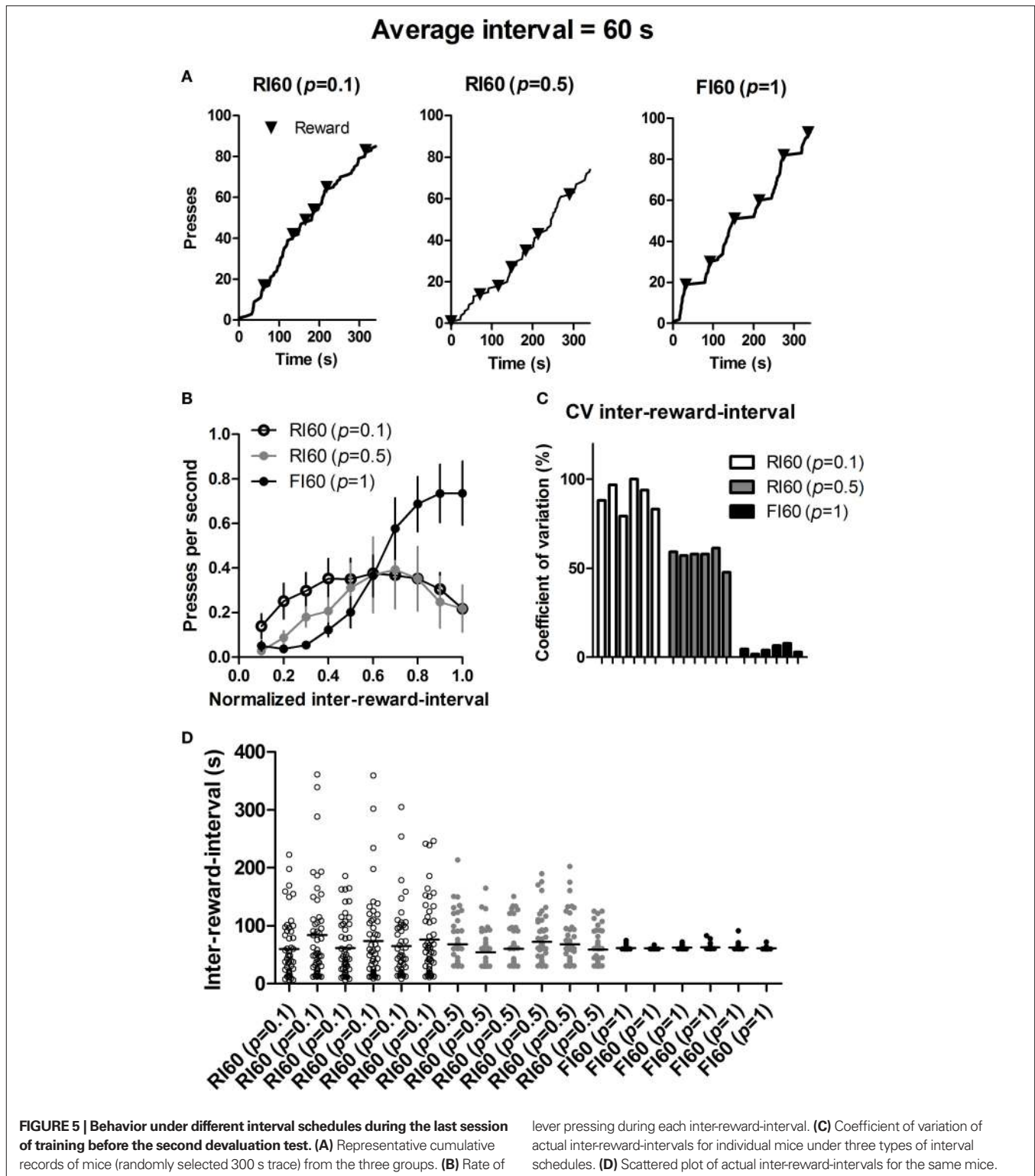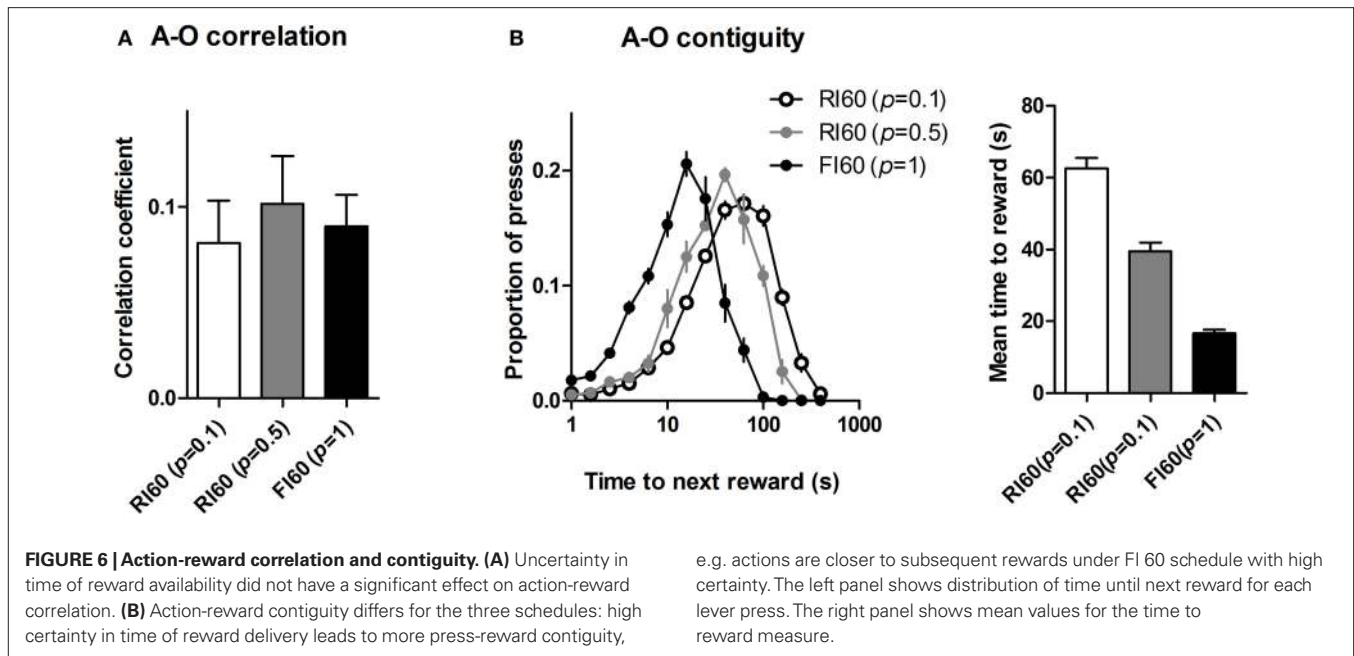

FIGURE 4 | Results from the second day of the omission test, expressed as a percentage of last training session. The left panel shows lever presses. The right panel shows the head entries into the food magazine.

1985). Despite the recent introduction of analytical behavioral assays in neuroscience, which permitted the study of neural implementation of operationally defined habitual behavior, the conditions that promote habit formation remain poorly characterized (Yin et al., 2004; Hilario et al., 2007; Yu et al., 2009).

FIGURE 5 | Behavior under different interval schedules during the last session of training before the second devaluation test. (A) Representative cumulative records of mice (randomly selected 300 s trace) from the three groups. **(B)** Rate of lever pressing during each inter-reward-interval. **(C)** Coefficient of variation of actual inter-reward-intervals for individual mice under three types of interval schedules. **(D)** Scattered plot of actual inter-reward-intervals for the same mice.

In this study we manipulated how 'random' the scheduled interval is, without changing the average rate of lever pressing, head entry, and reward (**Figure 2**). This manipulation significantly affected the pattern of lever pressing (**Figure 5**), the sensitivity of the behavior to changes in outcome value (specific satiety outcome devaluation, **Figure 3**), and sensitivity of the behavior to changes in the instrumental contingency (omission test, **Figure 4**). As our results show, uncertainty about the time of reward availability can promote habit formation, possibly by generating specific behavioral patterns with low action-reward contiguity.

FIGURE 6 | Action-reward correlation and contiguity. (A) Uncertainty in time of reward availability did not have a significant effect on action-reward correlation. **(B)** Action-reward contiguity differs for the three schedules: high certainty in time of reward delivery leads to more press-reward contiguity, e.g. actions are closer to subsequent rewards under FI 60 schedule with high certainty. The left panel shows distribution of time until next reward for each lever press. The right panel shows mean values for the time to reward measure.

On the early devaluation test, conducted after limited instrumental training (two sessions of 60-s interval schedules) all three groups were equally sensitive to the reduction in outcome value (**Figure 3**). With additional training (four additional sessions under the same schedule), however, a late devaluation test showed that only the FI group (low uncertainty) reduced lever pressing following specific satiety treatment. On the omission test, in which the reward is delivered automatically in the absence of lever pressing but canceled by lever pressing (Yin et al., 2006), the FI group also showed more sensitivity to the reversal in instrumental contingency (**Figure 4**).

Despite similar global feedback functions and average rates of reinforcement, FI and RI schedules are known to generate different patterns of behavior (Ferster and Skinner, 1957). For example, after extensive training FI schedules can produce a 'scalloping' pattern in the lever pressing, with prominent pauses immediately after reinforcement, and accelerating pressing as the end of the specified interval is approached; RI schedules, by contrast, maintains a much more constant rate of lever pressing (**Figure 5**). Under FI, the time period immediately after reinforcement signals no reward availability. Thus mice, just like other species previously studied (Gibbon et al., 1984), can predict the approximate time of reward availability, as indicated by their rate of lever pressing during each interval (**Figure 5B**).

Interval schedules in general have been thought to promote habit formation. It was previously proposed that the schedule differences in outcome devaluation could be explained by their feedback functions (Dickinson, 1989). According to this view, the molar or global correlation between the rate of action and the rate of outcome is the chief determinant of how 'goal-directed' the action is. The more the animal experiences such a contingency, the stronger the action-outcome representation and consequently the more sensitive behavior will be to manipulations of the outcome value and instrumental contingency. Because the overall feedback functions do not differ between FI and RI schedules, despite the difference in sensitivity of performance to outcome devaluation, a simple explanation in terms of the feedback function fails at the 'molar' level. But this does not mean that the experienced behavior-reward contingency is irrelevant (Dickinson, 1989). Because the different interval schedules we used do not differ much in terms of their global feedback functions, but produce strikingly distinct patterns of behavior, a more 'molecular' explanation of how RI schedules promote habit formation may be needed. However, the correlation between lever pressing and reward delivery was comparable across the three groups, suggesting that action-reward correlation was not responsible for the differences in sensitivity to devaluation and omission (**Figure 6A**).

A simple measure that does distinguish the behaviors generated by the different interval schedules we used is action-reward contiguity – the time between each lever press and the consequent reward, as illustrated in **Figure 6B**. The time between lever press and reward was on average much shorter under the FI schedule. Uncertainty in the time of reward availability resulted in more presses that are temporally far away from the subsequent reward.

Much evidence in the literature suggests a critical role for simple contiguity in instrumental learning and in determining reported causal efficacy of intentional actions in humans (Dickinson, 1994). Of course non-contiguous rewards presented in the absence of actions (i.e. instrumental contingency degradation) can also reduce instrumental performance even when action-reward contiguity is held constant (Shanks and Dickinson, 1991). But the presentation of non-contiguous reward engages additional mechanisms like contextual Pavlovian conditioning, which can produce behavior that competes with instrumental performance. In the absence of free rewards, however, action-reward contiguity is a major determinant of perceived causal efficacy of actions. Manipulations like the imposition of omission contingency effectively force a delay

between action and reward, thus reducing temporal contiguity. Therefore, a parsimonious explanation of our results is that the high action-reward contiguity in FI-generated lever pressing is responsible for greater goal-directedness in the behavior, as measured by devaluation and omission, and that habit formation under RI schedules is due to reduced action-reward contiguity experienced by the mice.

UNCERTAINTY

It is worth noting that in this study we did not manipulate action-reward contiguity. We manipulated the uncertainty in the time of reward availability. Increasing delay between action and outcome by itself is known to impair instrumental learning and performance (Dickinson, 1994). A direct and uniform manipulation of the action-reward delay *per se* is actually not expected to generate comparable rates of lever pressing. Given the analysis above, the question is how uncertainty in the time of reward availability can reliably produce predictable patterns of behavior. The influence of uncertainty on behavioral policy has not been examined extensively, though the concept of uncertainty has in recent years attracted much attention in neuroscience (Daw et al., 2005). One commonly used definition is similar to the concept of risk made popular by Knight (Knight, 1921). For example, in a Pavlovian conditioning experiment, a reward is delivered with a certain probability following a stimulus, independent of behavior (Fiorillo et al., 2003). Under these conditions, uncertainty, like entropy in information theory, is maximal when the probability of the reward given a stimulus is 50%, as in a fair coin toss (the least amount of information about the reward given a stimulus). Though mathematically convenient, this type of uncertainty is not very common in the biological world.

Rather different is the uncertainty in the time of reward availability in this study. As mentioned above, interval schedules model naturally depleting resources. A food may become available at regular intervals, but how 'regular' the intervals are can vary, being affected by many factors. Above all, there is an action requirement. When a fruit ripens, the animal does not necessarily possess perfect knowledge of its availability. Such information can only be discovered by actions. Nor, for that matter, is the food automatically delivered into the animal's mouth, as in laboratory experiments using Pavlovian conditioning procedures (Fiorillo et al., 2003). Purely Pavlovian responses, which are independent of action-outcome contingencies, are of limited utility in gathering information and finding rewards (Balleine and Dickinson, 1998). Hence the inadequacy of the purely passive Pavlovian interpretation of uncertainty often found in the economics literature, an interpretation that leaves out any role for actions.

Whatever the mouse experiences or does in the present study is not controlled directly by the experimenter, because it is up to the mouse to press the lever. If it does not press, no uncertainty can be experienced. But the mouse does behave predictably, in order to control food intake, because it is hungry. Thus the predictable patterns of behavior stem from internal reference signals for food, if we view the hungry animal simply as a control system for food rewards. When the time of reward availability

is highly variable under the RI 60 ($p = 0.1$) schedule, it presses quite constantly during the inter-reward-interval, a characteristic pattern of behavior under RI schedules (**Figure 5**). Such a policy ensures that any reward is collected as soon as it becomes available. The delay between the time of reward availability and the time of first press afterwards is, on average, simply determined by the rate of pressing under RI schedules (Staddon, 2001). One consequence of such a policy is reduced contiguity between lever pressing and actual delivery of the reward, as mentioned above (**Figure 6**). By contrast, when the uncertainty is low as in the FI schedule, mice can easily time the interval, increasing the rate of lever pressing as the scheduled time of reward availability approaches. Consequently, the contiguity between action and reward is higher in FI schedules. Therefore manipulations of 'temporal' uncertainty produce distinct behavioral patterns from animals seeking to maximize the rate of food intake. That such behavioral policies lead to major differences in experienced action-reward contiguity explains why different interval schedules can differ in their capacity to promote habit formation.

A useful analogy may be found in the behavior of email checking in humans. Suppose you would like to read emails from someone important to you as soon as they are sent, but this person has a rather unpredictable pattern of writing emails (RI schedule, high uncertainty). How do you minimize the delay between the time the email is sent and the time you read it? You check your email constantly. Of course you do not, unfortunately, have control over when your favorite emails are available, but you do, fortunately or unfortunately, have control over how soon you read them after they are sent. But herein lies the paradox: the more frequently you check your email, the shorter the delay between email availability and email reading, but at the same time the more often your checking behavior will be unrewarded by the discovery of a new email from your favorite person. That is to say, as you reduce the delay between reward availability (email sent) and reward collection (email read) you also increase the average delay between action (checking) and reward (reading).

SUMMARY AND NEUROBIOLOGICAL IMPLICATIONS

In short, our results suggest that the reduced sensitivity to outcome devaluation and omission under RI schedules can be most parsimoniously explained by the reduced action-reward contiguity in behavior generated by such schedules. This is a simple consequence of the behavioral policy pursued by animals to maximize the rate of reward (minimizing the delay between scheduled availability and actual receipt), without knowing exactly when the reward will be available.

Whether the generation of actions that are not contiguous with rewards will promote habit formation remains to be tested; nor is it clear from present results whether reduced action-reward contiguity is a *sufficient* explanation. A clear and testable prediction is that, in addition to uncertainty about reward availability, any experimental manipulation that results in reduced action-reward contiguity could promote habit formation. Such a possibility certainly has significant neurobiological implications. Considerable evidence shows that instrumental learning and performance depend on the cortico-basal ganglia networks, in particular the striatum, which is the main input nucleus and the target of massive dopaminergic

projections from the midbrain (Wickens et al., 2007; Yin et al., 2008). If rewards are typically associated with some reinforcement signals often attributed to a neuromodulator like dopamine released at the time of reward (Miller, 1981), then the neural activity associated with each action may be followed by different amounts of dopamine depending on temporal contiguity. Naturally, this is only one possibility among many. What actually occurs can only be revealed by direct measurements of dopamine release during behavior under interval schedules.

ACKNOWLEDGMENTS

REFERENCES

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol.* 33b, 109–122.

Adams, C. D., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.* 33, 109–122.

Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419.

Baum, W. M. (1973). The correlation-based law of effect. *J. Exp. Anal. Behav.* 20, 137–153.

Catania, A. C., and Reynolds, G. S. (1968). A quantitative analysis of the responding maintained by interval schedules of reinforcement. *J. Exp. Anal. Behav.* 11(Suppl), 327–383.

Colwill, R. M., and Rescorla, R. A. (1986). "Associative structures in instrumental learning," in *The Psychology of Learning and Motivation*, ed. G. Bower (New York: Academic Press), 55–104.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711.

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 308, 67–78.

Dickinson, A. (1989). "Expectancy theory in animal conditioning," in *Contemporary Learning Theories*, eds S. B. Klein and R. R. Mowrer (Hillsdale, NJ: Lawrence Erlbaum Associates), 279–308.

Dickinson, A. (1994). "Instrumental conditioning," in *Animal Learning and Cognition*, ed. N. J. Mackintosh (Orlando: Academic), 45–79.

Dickinson, A., Nicholas, D. J., and Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Q. J. Exp. Psychol. B* 35, 35–51.

Farmer, J. (1963). Properties of behavior under random interval reinforcement schedules. *J. Exp. Anal. Behav.* 6, 607–616.

Ferster, C., and Skinner, B. F. (1957). *Schedules of Reinforcement*. New York: Appleton Century.

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299, 1898–1902.

Gibbon, J., Church, R. M., and Meck, W. H. (1984). Scalar timing in memory. *Ann. N. Y. Acad. Sci.* 423, 52–77.

Hilario, M. R. F., Clouse, E., Yin, H. H., and Costa, R. M. (2007). Endocannabinoid signaling is critical for habit formation. *Front. Integr. Neurosci.* 1:6. doi: 10.3389/neuro.07.006.2007.

Knight, F. (1921). *Risk, Uncertainty, and Profit*. Boston: Hart, Schaffner, and Marx.

Miller, R. (1981). *Meaning and Purpose in the Intact Brain*. New York: Oxford University Press.

Shanks, D. R., and Dickinson, A. (1991). Instrumental judgment and performance under variations in action-outcome contingency and contiguity. *Mem. Cognit.* 19, 353–360.

Staddon, J. E. R. (2001). *Adaptive Dynamics: The Theoretical Analysis of Behavior*. Cambridge: MIT Press.

Tanaka, S. C., Balleine, B. W., and O'Doherty, J. P. (2008). Calculating consequences: brain systems that encode the causal effects of actions. *J. Neurosci.* 28, 6750–6755.

Wickens, J. R., Horvitz, J. C., Costa, R. M., and Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. *J. Neurosci.* 27, 8181–8183.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* 166, 189–196.

Yin, H. H., Ostlund, S. B., and Balleine, B. W. (2008). Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur. J. Neurosci.* 28, 1437–1448.

Yu, C., Gupta, J., Chen, J. F., and Yin, H. H. (2009). Genetic deletion of A2A adenosine receptors in the striatum selectively impairs habit formation. *J. Neurosci.* 29, 15100–15103.