

University of Vermont

ScholarWorks @ UVM

UVM Honors College Senior Theses

Undergraduate Theses

2017

Neuroimaging of Habit-based vs. Goal-directed behavior in Instrumental Learning

Michael Jacob Lawler
University of Vermont

Follow this and additional works at: <https://scholarworks.uvm.edu/hcoltheses>

Recommended Citation

Lawler, Michael Jacob, "Neuroimaging of Habit-based vs. Goal-directed behavior in Instrumental Learning" (2017). *UVM Honors College Senior Theses*. 156.
<https://scholarworks.uvm.edu/hcoltheses/156>

This Honors College Thesis is brought to you for free and open access by the Undergraduate Theses at ScholarWorks @ UVM. It has been accepted for inclusion in UVM Honors College Senior Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.

Neuroimaging of Habit-based vs. Goal-directed behavior in Instrumental Learning

Michael J. Lawler
Department of Neuroscience
University of Vermont

Supervised by
Dr. Scott Mackey

College of Arts and Sciences Sponsors

Dr. Mark Bouton (currently on Sabbatical)

&

Dr. John Green

Abstract

Addiction has been proposed to result from an overreliance on the habit-based and goal-directed controllers of behavior; however, few data exist to simultaneously support both behavioral and neuroanatomical aspects of this theory in humans. Here, we clarify the locations of the homologous structures controlling behavior in the human brain to those studied in animal models. The study included two parts. 1) The first part established in a behavioral experiment that the devaluation video in the present paradigm was able to influence instrumental behavior. Using a 3-session instrumental learning task to examine behavior, we examined 78 participants, aged 18-35. A significant difference in the change in response rate immediately before and after devaluation was found between the 2 groups viewing worms in devaluation compared to the group not viewing worms. There was a significant difference in change in liking immediately before and after devaluation between the three conditions, as well as in the change in liking, hunger, and response rate between the paired and empty bowl unpaired conditions. There was a significant correlation between snack liking pre-session 3 and response rate in session 3, as well as between pre-extinction snack liking and response rate in the start of extinction.

2) The second part of the study used the same 3-session training paradigm over 3-days, with fMRI on the third day to measure neural activity during this same instrumental learning task. Although the results are preliminary (N=10), these show that the comparable regions of the human brain are involved in goal-directed and habit-based control of behavior, with a perfect negative Spearman correlation of mean vmPFC activity at the end of training and the change in responding from immediately before to immediately after devaluation.

Three of the 10 subjects were addicted smokers, which is insufficient data to determine whether they were less sensitive to reward devaluation and whether they relied more heavily on brain structures associated with habit-based controllers of behavior. However, understanding the relationship between habit-based and goal-directed controllers of behavior and their role in addiction and clarifying the human brain structures responsible for these systems can lead to the development of therapies for addiction.

Introduction

An instrumental behavior is defined as any deliberate response that is reinforced when a desirable outcome that is contingent on that response occurs; this is often studied by training subjects to press a button or pull a lever for a rewarding outcome (Skinner, 1938). Although extensive research has identified a network of brain areas involved in instrumental behavior in animals, the location of the comparable areas in the human brain has been less well studied (Balleine & Dickinson, 1998a).

Two anatomically distinct controllers of instrumental behavior, the goal-directed and habit-based systems, have been identified in rodents. Goal-directed behaviors are deliberate and sensitive to the changes in the value of the outcomes subsequent to initial learning (Balleine & O'Doherty, 2010). Unlike goal-directed behaviors, the habit-based controller is not immediately affected by changes in the value of an outcome (Balleine & O'Doherty, 2010). The habit-based controller is more likely to control behavior after overtraining, when under a heavy cognitive load, or during stress (Balleine & O'Doherty, 2010).

The behavioral differences between controllers are illustrated by studies in which rats learn that the pressing of a lever earns a specific reward that is subsequently devalued (Balleine & Dickinson, 1998a; Balleine & Dickinson, 1998b). After initial learning, the outcome value can be changed if the reward is, for example, paired with LiCl to induce illness (Adams, 1982). If the behavior is controlled by the goal-directed system, the rats will press the button less after devaluation (Adams, 1982). An extinction period is one in which instrumental behavior is not reinforced, which leads to a termination of said behavior; an extinction period after devaluation can allow for an examination of the effects of devaluation on said behavior (Skinner, 1979). If instrumental responding persists even after devaluation, it is referred to as a habit regulated by the habit-based system (Balleine & O'Doherty, 2010). In the present example, rats would continue to press the button even though it has been devalued, and the rate of extinction would be slower than if the behavior was goal-directed (Adams,

Literature Review

Instrumental behavior is defined as any behavior that is performed in order to reach a goal, receive a reward, or remove an undesirable punishment (Skinner, 1938; Balleine & Dickinson, 1998a; Balleine & Dickinson, 1998b; Balleine & O'Doherty, 2010). Two neural systems that have been studied extensively in animals in regards to instrumental behavior are those that are responsible for habit-based and for goal-directed behaviors (Balleine & Dickinson, 1998a; Balleine & Dickinson, 1998b). These systems are distinct in the effects each has on controlling instrumental behavior (Balleine & Dickinson, 1998a). Goal-directed behaviors are purposeful and allow for flexible adjustment to rapid change in the perceived value of the behavioral outcome (Balleine & O'Doherty, 2010). Habit-based control, on the other hand, is fast, automatic, and more likely to control behavior after overtraining, when under a heavy cognitive load, or during stress (Balleine & O'Doherty, 2010). While habits allow for quicker and more efficient behavior under certain conditions, they are not affected by changes in the value of an outcome subsequent to initial learning (Balleine & Dickinson, 1998a; Balleine & Dickinson, 1998b; Balleine & O'Doherty, 2010).

These systems are also distinct anatomically. In rats, the goal-directed system relies on the prelimbic cortex and the dorsomedial striatum (Balleine & Dickinson, 1998), while anatomically comparable structures in the human brain are the ventromedial prefrontal cortex and the anterior caudate nucleus of the basal ganglia (Valentin, Dickinson, & O'Doherty, 2007; Liljeholm, Tricomi, O'Doherty, & Balleine, 2011). The habit-based system relies more heavily on the dorsolateral striatum in rats (Balleine & Dickinson, 1998). The analogous substrate in humans is thought to be located in the dorsolateral posterior putamen, a region within the striatum (Tricomi, Balleine, & O'Doherty, 2009).

Highlighting the distinctions between habit-based and goal-directed controllers of behavior and their anatomical substrates may be an important component in understanding the mechanisms of

addiction (Sjoerds, 2013). Experiments in animal models have supported the hypothesis that addiction arises from the overreliance on habit-based behavior in tandem with underutilization of goal-directed behavior (Balleine et al., 2014). However, relatively little neuroimaging research has been done to test how findings on instrumental learning in animals apply to the human brain (Valentin et al., 2007; Tricomi et al., 2009; McNamee et al., 2015). Using fMRI, Tricomi et al. (2009) found that, as training progressed, the dorsolateral striatum was activated more at the onset of task blocks than at the onset of rest blocks, providing preliminary evidence in the human brain of a shift from goal-directed to habit-based control of behavior as overtraining occurs. Sjoerds et al. (2013) provided further evidence for the role of habit-based learning in human addiction in their neuroimaging study on habit learning in alcohol-dependent subjects. In this study, increased dorsolateral striatum activity was observed in the habit-based learning condition in alcohol-dependent subjects than in healthy controls, while greater ventromedial prefrontal cortex activity was observed in the goal-directed learning condition in healthy controls than in alcohol-dependent subjects (Sjoerds et al., 2013). These findings support the theory that subjects with substance dependence disorders have a higher propensity for habit-based control of behavior than healthy controls (Sjoerds et al., 2013).

The problem with this interpretation of the Sjoerds study is that there is a body of literature suggesting that substance dependence is associated with decreased cognitive ability (e.g. Mintzer & Stitzer, 2002; Verdejo-García, Bechara, Recknor, & Pérez-García, 2006; Scott et al., 2007). As a result, it is possible that the Sjoerds results could be explained by the lower cognitive capacity of the alcoholics, due to their task's complicated instructions and high workload on working memory (Mintzer & Stitzer, 2002; Verdejo-García, Bechara, Recknor, & Pérez-García, 2006; Scott et al., 2007; Sjoerds et al. 2013). Therefore, it is necessary to conduct a study that examines the differences between goal-directed and habit-based behaviors that is not confounded by the issues of cognitive ability.

The present study used a simple task that was not confounded by cognitive ability (Tricomi et

al, 2009). This experimental paradigm was based directly on the previous experiments performed in animals, so there is strong support for its simplicity, as well as its effectiveness in examining these two controllers of behavior in animals (Balleine & Dickinson, 1998a; Balleine & Dickinson, 1998b; Balleine & O'Doherty, 2010). Moreover, a similar design was used by Tricomi et al. (2009) to study schizophrenics, so its efficacy in humans has been demonstrated as well.

My study will help further clarify the locations of the homologous structures of the habit-based and goal-directed systems in the human brain to those studied in animal models. Ultimately, when the study is complete, it will provide evidence of the possible connections between the habit-based and goal-directed controllers of behavior and addiction. It will tackle a similar problem to the study on alcohol-dependent subjects performed by Sjoerds et al. (2013), but will have an approach more closely related to that used by Tricomi et al. (2009), and will develop a paradigm that will be able to avoid the potentially confounding effects of substance dependence on cognitive ability. For the behavioral pilot data, we hypothesized that there would be a greater gross decrease and % decrease in response rate between the last session of training and the start of extinction 1) between the three groups; 2) between the paired group compared to the unpaired groups; and 3) between the groups that saw worms compared to the group that did not. Additionally, we predicted that liking would predict the rate of responding after devaluation.

Objectives

Part 1. The original paradigm included a 3-day study design in which task training occurred over three consecutive days, with the fMRI scan occurring only during the training, devaluation, and extinction on the third day. Because it is difficult to schedule a participant for three consecutive days at the same time of day and because having three separate study visits increases the chances of study dropout, a 1-day, 3-session paradigm was proposed. This 1-day paradigm was not used for the fMRI subjects, because the multi-day paradigm allows us to identify and exclude problematic participants

before performing an expensive MRI scan. However, in order to assess the potential efficacy of this new paradigm for the future, pilot data was obtained. This pilot paradigm consisted of a single hour-long study visit of three consecutive training sessions without an fMRI scan. This aimed to test whether or not the behavioral principles would translate to a single day. More importantly, this pilot examined whether the change of behavior during the extinction phase of the paradigm was due specifically to the content of the paired variable or could be explained by some other factor (e.g. fatigue). Obtaining this pilot behavioral data constituted a large portion of the research conducted in the present study.

Part 2. The primary purpose of the second portion was to determine whether human brain regions involved in habit-based and goal-directed behavior are comparable to brain regions identified in rodent studies. Specifically, more ventromedial prefrontal cortex (vmPFC) activity was expected if the observed behavior were more goal-directed, whereas more putamen activity was expected if the observed behavior were more habit based. Testing this hypothesis requires an experimental paradigm that is effective in demonstrating the differences in habit-based and goal-directed controllers of behavior in human subjects.

This experiment is part of a larger project being conducted in the lab of Dr. Scott Mackey in collaboration with Dr. Hugh Garavan and Dr. Eric Thraillkill, which seeks to compare differences in habit-based responding between smokers and nonsmokers. The ultimate goal of this project is to better understand the relationship between habit-based and goal-directed controllers of behavior and their role in addiction, to clarify the human brain structures that are analogous to those defined in animals, and to develop therapies for addiction based on these findings.

Methods

Participants

In the behavioral pilot there were 116 participants ages 18-35 recruited from both the local community and from UVM through the Sona Systems Psychological Science Participant System. Each

participant had no history of diabetes, was not currently dieting, and had no visual or physical impairments that would interfere with viewing or using a computer screen and keyboard. Of the 116 participants, 38 were excluded for a final total of $N=78$. Exclusion was due to a number of reasons, including failure to understand or comply with task expectations, incomplete task participation, or insufficient motivation for the task indicated by fewer than 20 responses during the third session of training. The 78 remaining participants were randomly assigned to three different experimental groups, each of which saw a different video after the third training session. The paired group saw a video of worms crawling on their chosen snack in the same bowl from which they ate snack during the task ($N=24$). One unpaired group saw a video of the same bowl from which they ate snack during the task with nothing in it ($N=23$). The second unpaired group saw a video of the same bowl from which they ate snack during the task with just worms in it, in order to attempt to control for the exposure to worms when not paired with the reward ($N=31$). Participants were compensated \$25 for participation if recruited from the community, and were compensated 1 credit to be used in a participating Psychological Science class of their choosing if they were recruited through Sona.

For the 3-day fMRI paradigm, 3 smokers and 11 nonsmokers were recruited. Imaging data on the first 3 subjects was collected prior to an upgrade in the scanner apparatus and software. The spatial resolution of these data was lower than the remainder of the study and so they are not reported here. One additional nonsmoker was excluded due to an incomplete scan for a total count of $N=10$. Individuals with comorbid substance abuse/dependence or who endorsed a lifetime history of Axis I disorders were excluded. All fMRI subjects viewed the same video, in which worms crawled over the rewards in the same type of bowl in which rewards were presented during training. Each subject was compensated \$75 for participation.

Instrumental Learning Task and Questionnaires

Scores on multiple personality measures suspected to predict preference of one controller of

behavior over the other were collected on the first day following consent, including the Barrett Impulsivity Scale (BIS), the Body Awareness Questionnaire (BAQ), the Toronto Alexithymia Scale (TAS), and the Big 5 Personality Traits. Nicotine and other drug use was characterized with the Fagerstrom Test for Nicotine Dependence (FTND) and substance use and abuse questionnaires adapted from the Composite International Diagnostic Interview (CIDI), with questions and scoring obtained from the PhenX Toolkit (<https://www.phenxtoolkit.org/>).

After giving informed consent and filling out the questionnaires, all subjects were asked to rate their hunger and liking of three reward options (apple chips, M&Ms, and potato chips) on a 7-point Likert scale. The reward with the highest liking was their reward for the entire study. Subjects in both the pilot and fMRI paradigms were trained to instrumentally respond for a reward of their choosing. Specifically, they were instructed that pressing a response key will tilt a virtual vending machine on a computer screen, with each tilt offering the opportunity to earn food rewards. Responding was voluntary and reinforced on a variable interval schedule using a fixed probability (i.e. 1000ms intervals with a 20% probability of reward), to most rapidly promote habit-based behavior.

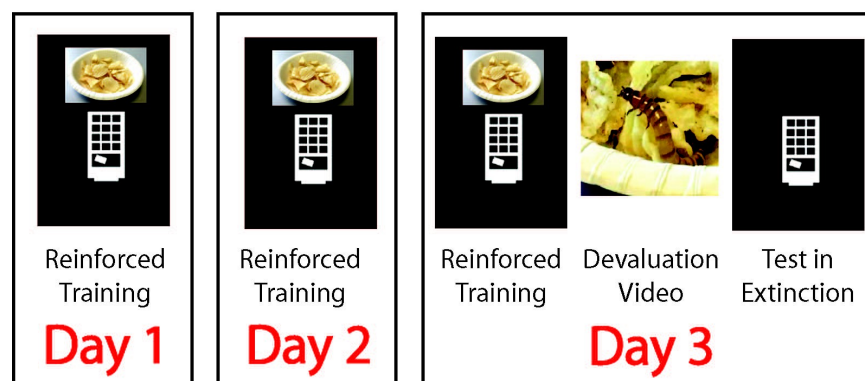


Figure 1. Task training occurred in 3 sessions over 3 days for the fMRI paradigm, and occurred in 3 sessions on 1 day for the behavioral pilot.

Rewards earned were indicated using a virtual image at a 2 images:1 reward ratio, and were delivered in a white paper bowl at the end of each of the 40-second blocks on day 1 of the three

consecutive training days (session 1 in the behavioral pilot), and at the end of the task on days 2 and 3 (sessions 2 and 3 in the behavioral pilot). Rewards were eaten before proceeding to the next block on day 1/session 1. Participants were asked to rate their hunger and liking of their chosen snack before and after training on days/sessions 1 and 2, before session 3, and after extinction. They were also asked to rate liking of their snack immediately following devaluation. Training on days 1 and 2 of the fMRI paradigm and all 3 sessions of the behavioral pilot were conducted in the Arnold and St. Joseph wings of the UHC campus of the UVM Medical Center.

Neuroimaging Measures

On day 3 of the fMRI paradigm, training was conducted in the UVM MRI Center for Biomedical Engineering at the UVM Medical Center in Burlington, Vermont. Neuroimaging using functional magnetic resonance imaging (fMRI) was conducted in a one-hour session using a 3.0 Tesla scanner. Baseline brain activity readings were obtained between each reinforced trial, and subjects with excessive head motion during scanning were excluded for image accuracy. After training on the third day, a 4-minute devaluation video was shown, in which worms crawl over the rewards in the same type of bowl in which rewards were presented during training. Morris and colleagues (2015) demonstrated the effectiveness of this devaluation video in decreasing instrumental responding, in order to examine behavior during extinction. In extinction, the same instrumental responding task was presented to subjects, but without reinforcement. The difference between the average response rate, hunger rating, and snack liking scores in the third session of training and those values from the first five blocks of extinction were obtained via subtraction and used in analyses.

Magnetic resonance images were collected on a 3.0 Tesla scanner (Philips 3T Achieva TX dStream MRI) equipped with 8 high bandwidth receivers that allowed for shorter read-out times and reduced signal distortions and ventromedial signal dropout. Structural MRI acquisition consisted of T1- and T2-weighted images at 0.8mm isotropic resolution, along with T2-FLAIR at 1.0mm isotropic

resolution. The high resolution T1- and T2-weighted images were well suited to the HCP processing streams, including the generation of cortical myelin maps [41, 58], while the T2-FLAIR provided excellent sensitivity to intra-cranial pathology. Task fMRI parameters were TR 800ms, TE 30ms, flip angle 52°, 2.4mm isotropic imaging resolution with a 216×216×144mm³ field of view using a multiband acceleration factor of 6 (60 slices, no gap). Structural and fMRI preprocessing used the pipelines developed as part of the Human Connectome Project (Smith et al., 2004). The HCP functional pipeline corrects for EPI spatial distortions using magnetic field maps, realigns volumes to account for subject motion and registers to the structural data. The functional data is then resampled into CIFTI grayordinate standard space. HCP task fMRI pipelines were used for first- and second-level analysis of task fMRI data. These pipelines are based on FSL FEAT (FMRIB's Expert Analysis Tool) and FLAME (FMRIB's Local Analysis of Mixed Effects), and incorporate high-pass filtering and application of general linear models (GLMs) (Tournier, et al., 2004; Tournier, Calamante, & Connelly, 2007). Finally, mean activity was extracted from regions of interest (ROIs) as defined by the Destrieux Atlas (Destrieux, Fischl, Dale, & Halgren, 2010).

Behavioral Analyses

Multiple analyses were performed with the behavioral pilot data. The first were ANOVAs between the 3 experimental conditions, examining the mean change and % change in response rate from the last 5 blocks of training to the first 5 blocks of extinction. These ANOVAs were then repeated with pre-extinction liking as a covariate. Next, independent samples t-tests of the change and % change in response rate from the end of training to the start of extinction between all of those who saw worms in the devaluation video (N=55) and those in the empty bowl unpaired group who did not see worms (N=23) were conducted. Also, t-tests of the change and % change in response rate immediately before and after devaluation were performed between paired (N=24) and combined unpaired (N=54) groups.

Repeated measures ANOVAs were also performed, examining hunger, liking, and response rate

during the first 3 sessions between the three conditions. ANOVAs were performed of the change in liking and hunger immediately before and after devaluation between 3 conditions, as well as independent t-tests of the change in hunger, liking, and response rate immediately before and after devaluation between the paired and combined unpaired conditions. The Pearson correlation between snack liking pre-session 3 and response rate in session 3 was examined, as well as the correlation between pre-extinction/post-devaluation liking and response rate during the start of extinction for each group separately, and all together.

For the MRI behavioral data, Independent-Samples-Mann-Whitney U tests of the change and % change in response rate immediately before and after devaluation between smokers and nonsmokers were performed. Finally, correlations of scores on a variety of different personality measures with change and % change in response rate immediately before and after devaluation were examined.

When incorporating the available fMRI scan data, a Spearman correlation between mean activity in the vmPFC at the end of training and percent change in responding between the last five blocks of training and the first five blocks of extinction was calculated.

Results

Behavioral Analyses – Response Rate

All significant p-values will be reported with * to denote significance at 1-tailed 0.05 alpha level. An ANOVA examining the change in response rate from the end of training to the start of extinction (the 5 blocks immediately before and after devaluation) including all three conditions was not significant ($F_{2,78} = 1.735$, $p = 0.183$) (Table 1). However, a t-test comparing just the paired (food and worms in a bowl) and unpaired (empty bowl) conditions was significant ($t_{45} = -1.776$, $p < 0.05$) (Figure 2 and Table 3). Table 1 also includes tests of the change in response rate expressed as percent.

A significant difference was not observed in the change of the rate of responding following devaluation for the empty bowl and worms only conditions ($t_{52} = -1.643$, $p > 0.05$; Figure 9 and Table

3), but a significant difference was observed in the change of the rate of responding following devaluation when the condition with the empty bowl was compared to the combined conditions with worms (i.e. the condition that paired worms and food & the condition with worms alone in bowl) ($t_{76} = -1.853, p < 0.05$) (Figure 4 and Table 3). This significance was not seen when examining mean differences between groups in the % change in response rate immediately before and after devaluation ($t_{76} = -1.037$) (Table 3).

Behavioral Analyses – Response Rate, Hunger Rating, and Snack Liking

To demonstrate that no group differences were present before devaluation, a repeated measures ANOVA over the 3 training sessions of training was performed and showed that the three conditions were not significantly different for hunger, liking, or response rate ($F = 0.116, df = 2, 78; F = 0.213, df = 2, 78; F = 0.084, df = 2, 78$ [respectively]; Table 4). An ANOVA comparing these values for the last 5 blocks of training (session 3) was also not significant for all three measures ($F = 0.139, df = 2, 78; F = 0.227, df = 2, 78; F = 0.056, df = 2, 78$ [respectively]; Table 4). The ANOVA between the 3 conditions for the mean change in hunger from the end of training to the end of extinction was not significant ($F = 2.069, df = 2, 78$; Table 5), but was significant at the 2-tailed 0.1 alpha level for the change in liking immediately before and after devaluation ($F = 2.908, df = 2, 78, p < 0.05$ (2-tailed); Figure 5 and Table 5). When only considering paired and empty bowl unpaired conditions, independent t-tests of change in liking, hunger, and response rate from immediately before to immediately after devaluation show significant differences between groups ($t_{45} = -2.164, p < 0.05$, Figure 6; $t_{45} = -1.863, p < 0.05$, Figure 7; $t_{45} = -1.776, p < 0.05$, Figure 8; Table 6).

Snack liking pre-session 3 and response rate in session 3 were significantly correlated ($r = 0.215, p < 0.05$; Table 7). Liking immediately following devaluation and response rate during the start of extinction were not significantly correlated at the 1-tailed 0.05 alpha level for any of the groups individually ($r = 0.288, r = 0.21, \text{ and } r = 0.109$, respectively) (Table 8), but were significantly

correlated across groups ($r = 0.227$, $p < 0.05$) (Table 8).

Behavioral Analyses – Response Rates and Personality in MRI Subjects

For the behavioral data obtained from the subjects of the fMRI paradigm, independent-Samples-Mann-Whitney U tests revealed no significant differences between smokers ($N=3$) and nonsmokers ($N=7$) for either change or % change in response rate immediately before and after devaluation ($p = 0.183$ and $p = 0.383$, respectively) (Table 9). Pearson correlations were also examined between scores on a variety of different personality measures and the change and % change in response rate immediately before and after devaluation, and only the Big 5 personality trait conscientiousness was shown to have a significant correlation ($r = -0.644$, $p < 0.05$) (Table 10).

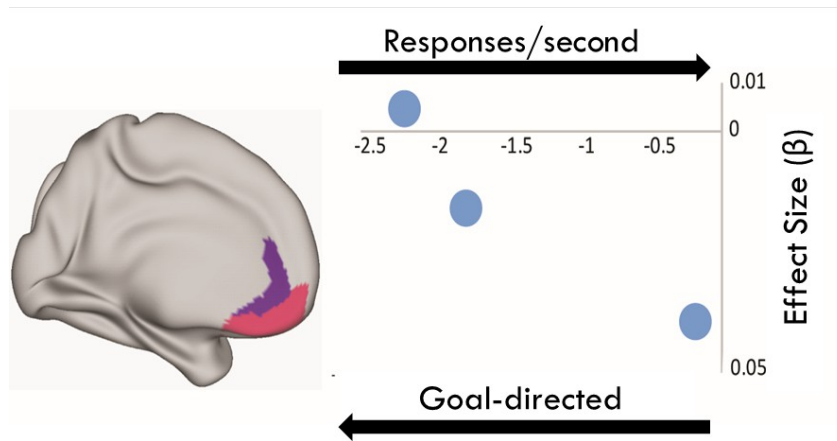
Neuroimaging Results

Figure 3. Left) Freesurfer ROIs occupying the vmPFC region; Right) Difference in responding during the 5 blocks immediately before and after devaluation plotted against MRI effect size. More vmPFC activity in individuals with the greatest change in responding after devaluation (i.e. goal-directed behavior).

Analysis of Freesurfer ROIs of the first three MRI participants (2 smokers and 1 nonsmoker) alongside behavioral data revealed a perfect negative Spearman correlation between the effect size of vmPFC activity and the difference in response rate from the 5 blocks immediately before and after devaluation ($r = -1$, $p < 0.01$). The data from the 7 remaining subjects have yet to be analyzed at this time, but analyses are continuing.

Figures and Tables

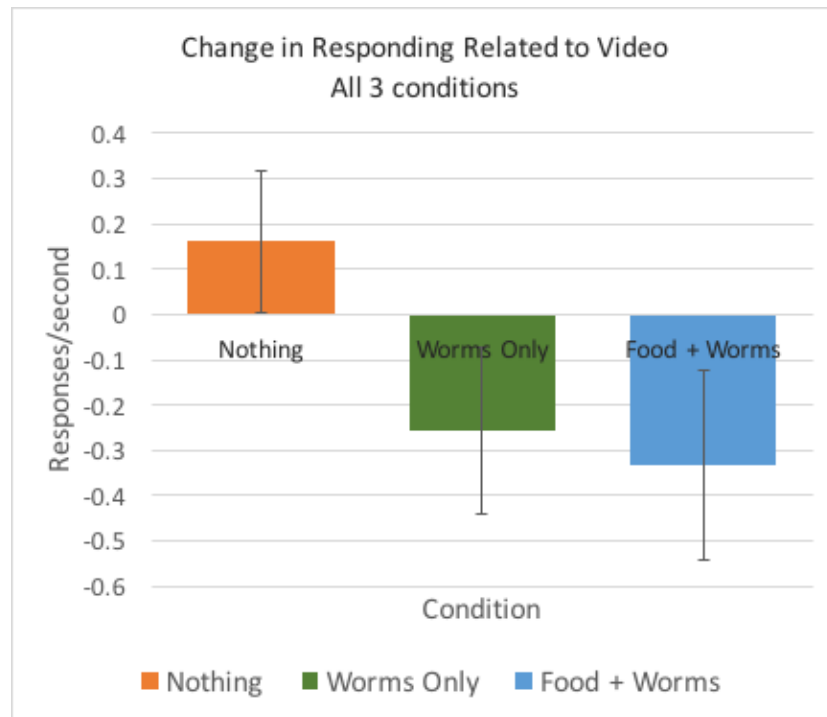


Figure 2. Means of change in responding immediately before and after devaluation related to the content of the devaluation video for all 3 conditions. ANOVA: $F_{2,78} = 1.735$, $p = 0.183$

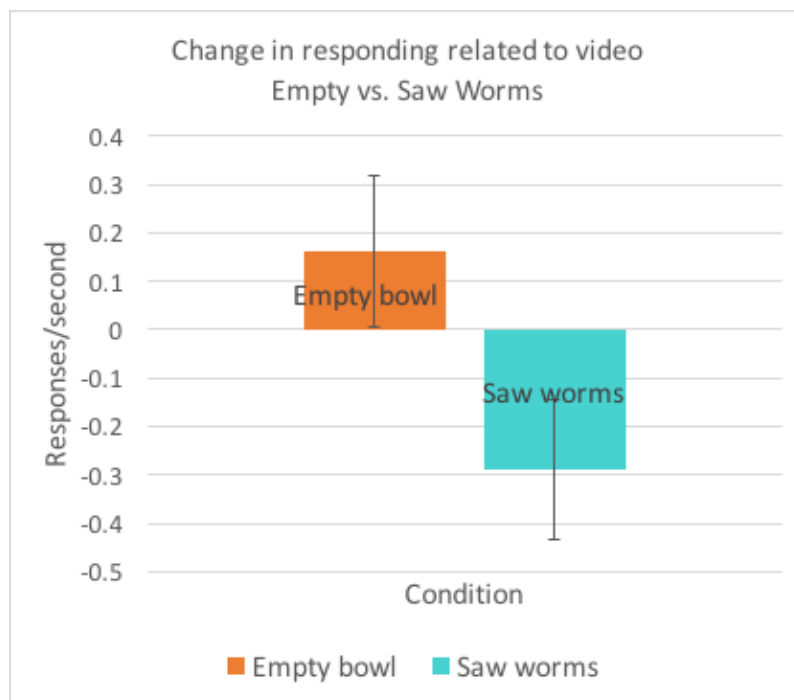


Figure 4. Change in responding immediately before and after devaluation related to the content of the devaluation video for those who saw worms (worms only and worms+food) vs. those who did not (empty bowl). Independent Samples t-test: $t_{76} = -1.853$, $p < 0.05$

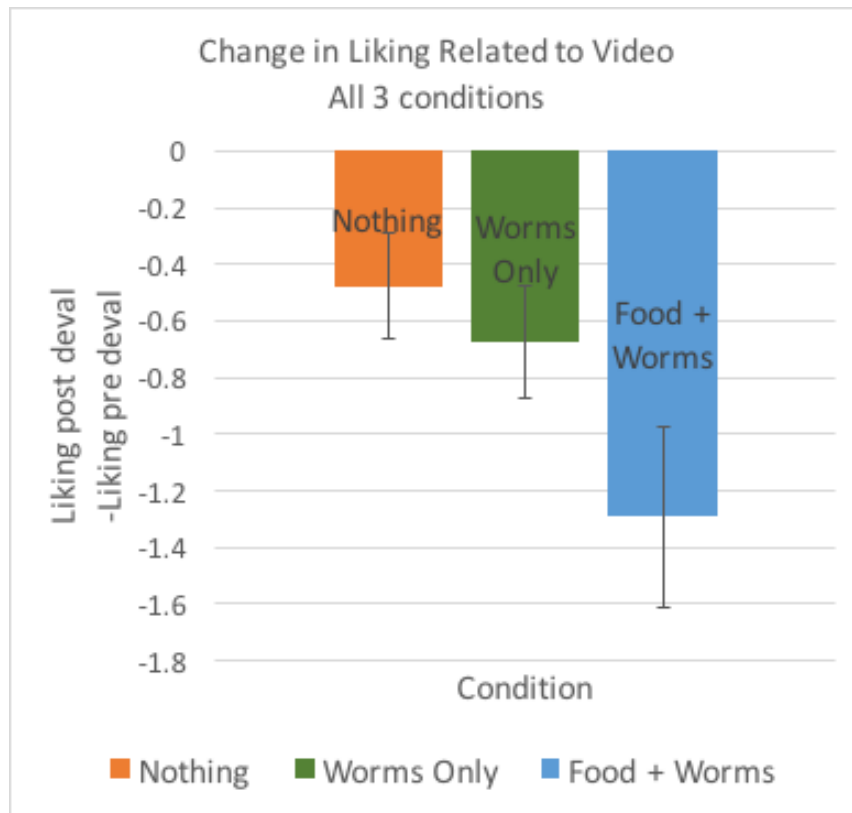


Figure 5. Change in liking immediately before and after devaluation related to the content of the devaluation video for all 3 conditions. ANOVA: $F = 2.908$, $df = 2, 78$, $p < 0.05$ (2-tailed)

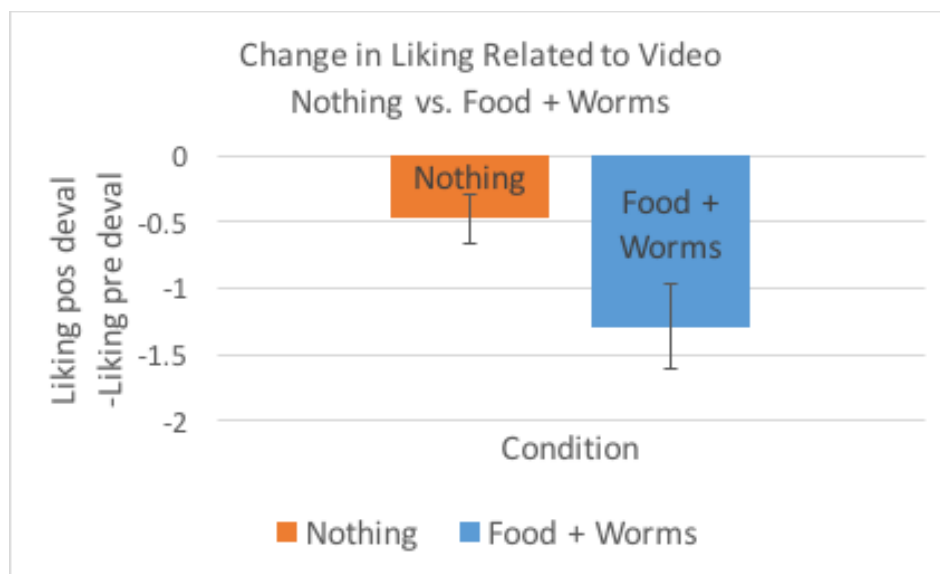


Figure 6. Change in liking immediately before and after devaluation related to the content of the devaluation video for paired (worms + food) vs. unpaired (empty bowl) conditions. Independent Samples t-test: $t_{45} = -2.164$, $p < 0.05$

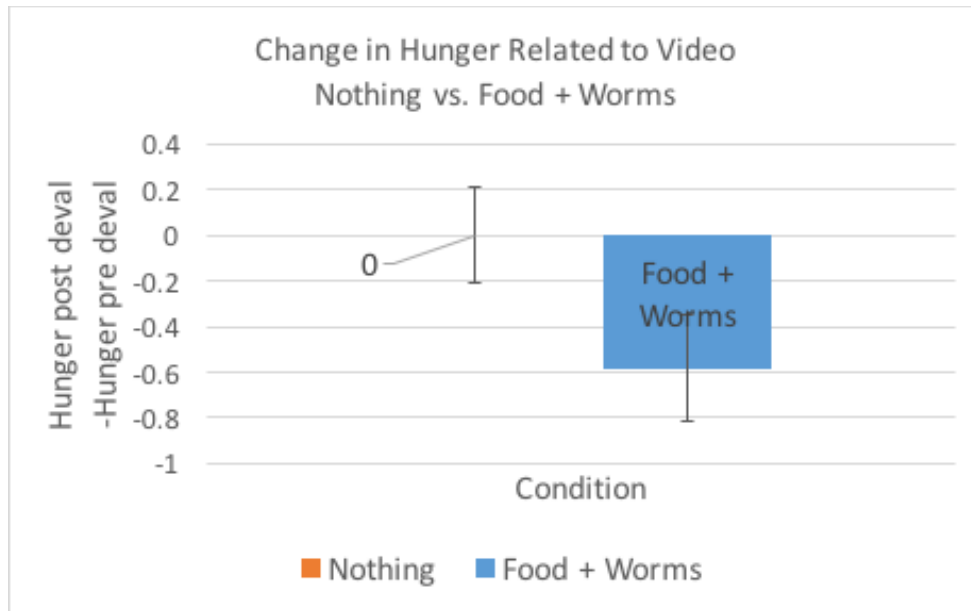


Figure 7. Change in hunger immediately before and after devaluation related to the content of the devaluation video for paired (worms + food) vs. unpaired (empty bowl) conditions. Independent Samples t-test: $t_{45} = -1.863$, $p < 0.05$

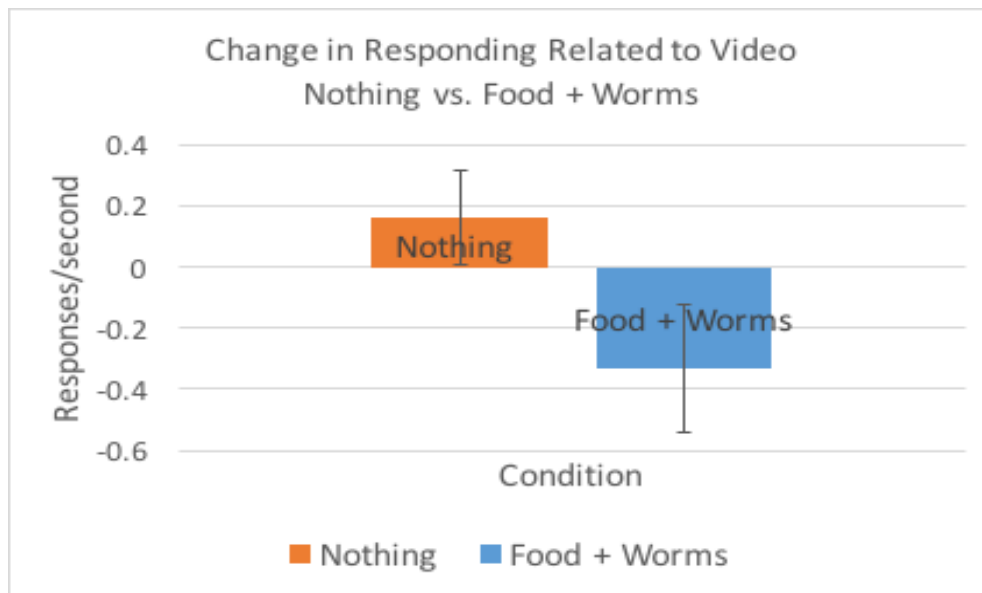


Figure 8. Change in responding immediately before and after devaluation related to the content of the devaluation video for paired (worms + food) vs. unpaired (empty bowl) conditions. Independent Samples t-test: $t_{45} = -1.776$, $p < 0.05$

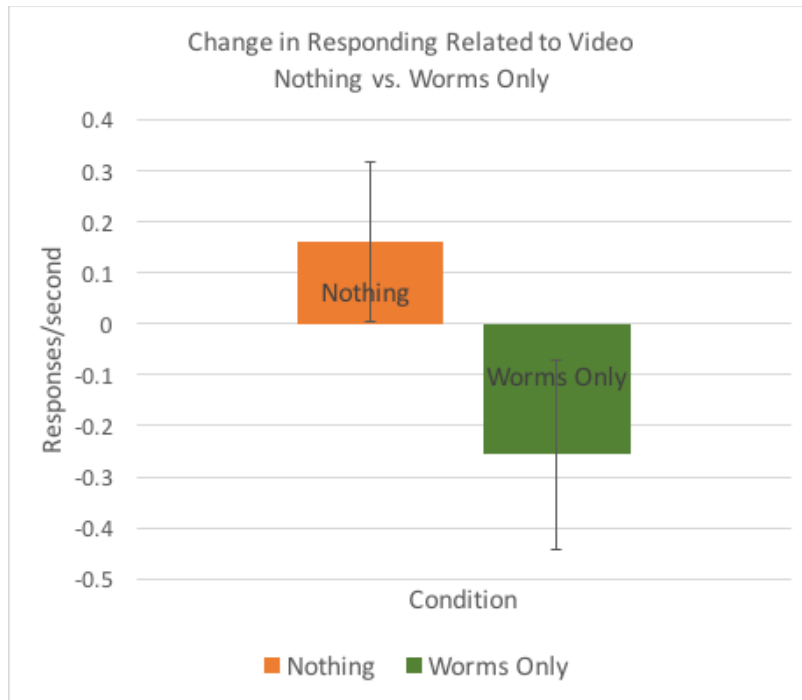


Figure 9. Change in responding immediately before and after devaluation related to the content of the devaluation video for unpaired (empty food) vs. unpaired (worms only) conditions. Independent Samples t-test: $t_{52} = -1.643$, $p > 0.05$

Table 1. ANOVAs between 3 experimental conditions of the mean change and % change in response rate from the last 5 blocks of training to the first 5 blocks of extinction. The three conditions were: 1) paired (worms + reward); 2) unpaired (empty bowl); and 3) unpaired (worms only).

Dependent Variable	F	df
Δ response rate from training! start of extinction	1.735	2,78
% Δ response rate from training! start of extinction	.828	2, 78

Table 2. ANOVAs between 3 experimental conditions of the mean change and % change in response rate from the last 5 blocks of training to the first 5 blocks of extinction with pre-extinction snack rating as a covariate.

Dependent Variable	F	df
Δ response rate from training! start of extinction	1.082	2, 78
% Δ response rate from training! start of extinction	.522	2, 78

Table 3. Independent samples t-tests of the Δ and % Δ in response rate from training! start of extinction between those who saw worms in the devaluation video (combined worms+food and worms-only groups) and those in the empty bowl group who did not see worms. Also t-tests of Δ response rate and % Δ response rate between paired and combined unpaired groups, as well as Δ response rate for the empty bowl vs. worms only groups.

Dependent	t	df
Δ response rate and Worms vs. No worms	-1.853 (worms<no worms)*	76
Δ response rate and Paired vs. Unpaired	-1.037 (paired<unpaired)	76
% Δ response rate and Worms vs. No worms	-1.292 (worms<no worms)	76
% Δ response rate and Paired vs. Unpaired	-0.635 (paired<unpaired)	76
Δ response rate and Empty bowl vs. Worms Only	-1.643 (worms only<unpaired)	52

Table 4. Comparisons of hunger, liking, and response rate between three conditions, with a repeated measures ANOVA over all 3 training sessions and ANOVA of session 3 alone

Repeated Measures ANOVA across all training			ANOVA of session 3	
Dependent Variable	F	df	F	df
Hunger	.116	2, 78	.139	2, 78
Liking	.213	2, 78	.227	2, 78
Response rate	.084	2, 78	.056	2, 78

Table 5. ANOVA of Δ liking and Δ hunger from end of training! extinction between 3 conditions.

Dependent Variable	F	df
Δ liking from end of training! start of extinction	2.908 *	2, 78
Δ hunger from end of training! end of extinction	2.069	2, 78

*. F statistic is significant at the 0.1 level (2-tailed).

Table 6. Independent samples t-tests of various variables immediately before and after devaluation between the paired and empty bowl unpaired conditions.

Dependent Variable	t	df
Δ liking from end of training! start of extinction	-2.164*	45
Δ hunger from end of training! end of extinction	-1.863*	45

Δ response rate from end of training! start of extinction	-1.776*	45
--	---------	----

*. Critical t-values significant at the 0.1 level (2-tailed).

Table 7. Correlation between snack liking (L) pre-session 3 and response rate in session 3.

	Pearson Correlation, r
Session 3 Resp./sec – Pre 3 L	.215*

*. Correlation is significant at the 0.05 level (1-tailed).

Table 8. Correlation between pre-extinction/post-devaluation liking and response rate during the start of extinction for each group separately and all together.

	Pearson Correlation, r
Group 1- Liking post deval. + resp. rate in start of extinction	0.288
Group 2- Liking post deval. + Resp. rate in start of extinction	0.21
Group 3- Liking post deval. + resp. rate in start of extinction	0.109
All groups- pre-ext. liking + resp. rate in start of extinction	0.227*

Table 9. Independent-Samples-Mann-Whitney U tests of the Δ and % Δ in response rate from training! start of extinction between smokers and nonsmokers in the fMRI paradigm.

Dependent	p
Δ training->start extinction smokers vs nonsmokers (fMRI)	p = .183
% Δ training->start extinction smokers vs nonsmokers (fMRI)	p = .383

Table 10. Correlations of scores on a variety of different personality measures with Δ and % Δ response rate immediately before and after devaluation.

Personality Measure	MRI Δ training->start extinction			MRI % Δ training->start extinction		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
Big 5- Extraversion	-0.4	0.252	10	-0.443	0.2	10
Big 5- Agreeableness	-0.042	0.908	10	-0.073	0.84	10
Big 5- Conscientiousness	-0.54	0.106	10	-.644*	0.044	10
Big 5- Neuroticism	0.237	0.51	10	0.27	0.452	10

Big 5- Openness	-0.337	0.34	10	-0.179	0.62	10
BAQ	0.64	0.122	7	0.462	0.296	7
TAS	0.183	0.696	7	0.358	0.43	7
BIS Total	-0.308	0.502	7	-0.358	0.43	7
BIS Attentional	0.334	0.464	7	-0.012	0.98	7
BIS Attention	0.447	0.316	7	0.205	0.66	7
BIS Cognitive Instability	0.091	0.846	7	-0.36	0.427	7
BIS Motor	-0.521	0.23	7	-0.261	0.572	7
BIS Motor	-0.649	0.114	7	-0.43	0.335	7
BIS Perseverance	-0.069	0.884	7	0.128	0.784	7
BIS Nonplanning	-0.329	0.472	7	-0.407	0.365	7
BIS Self-Control	-0.087	0.852	7	-0.055	0.907	7
BIS Cognitive Complexity	-0.376	0.406	7	-0.518	0.234	7

Discussion

In the present study, we examined the relationship between the habit-based and goal-directed controllers of behavior through a behavioral task and neuroimaging. Using statistical analyses of behavioral data, we highlighted some behavioral evidence supporting a distinction between habit-based and goal-directed controllers of behavior and supporting the devaluation effect of the task paradigm. Using fMRI, we demonstrated that activity in the vmPFC was related to greater goal-directed control of behavior. These findings support previous literature showing the distinctions between these two behavior controllers, and also offer new insight into the intricacies of examining instrumental learning in humans.

As expected, there was a significant difference between the condition that paired worms with food in a bowl and the condition that showed an empty bowl alone (Figure 2). This demonstrates that the change in behavior is due to the content of the devaluation video and not due to other factors related to the experimental context, such as fatigue. The condition with worms in an empty bowl was intermediate between these two.

One possible explanation for this is that, while they could not have a direct outcome devaluation effect, the worms seen in the video for that group devalued the empty bowl – that is to say, the context was devalued (Bouton & Bolles, 1979). Context dependent learning has been demonstrated in many studies; for example, context-dependent drug anticipation is thought to contribute to drug overdose when a frequent drug user takes a drug in an unfamiliar context (Siegel, 1999). To test this hypothesis, those who saw worms from the paired group and those in the worms only unpaired groups were combined into one group (N=55) and the change in response rate from immediately before and after devaluation was compared to the empty bowl unpaired group (N=23). This independent samples t-test showed a significant difference, supporting the idea that the presence of worms in the video decreased responding. Showing that the presence of worms in the context lead to decreased responding regardless of the presence of the reinforcer is supported by literature on contextual control of instrumental learning group (Bouton, Todd, Vurbic, & Winterbauer, 2011; Thrailkill & Bouton, 2014). This suggests that those who decreased responding in the worms only condition exhibited goal-directed behavior that was sensitive to context devaluation.

Even with this evidence, other explanations are possible; for example, it is possible that the worms in the video caused a disgust reaction, thereby reducing hunger and subsequently responding (Lovibond, 1981). The insignificant differences between groups in hunger, liking, and response rate found by the repeated measures ANOVA indicates that there were no differences in these three measures between the three conditions prior to devaluation and extinction, and that therefore differences seen in these measures after devaluation and extinction can be attributed to the condition. However, the ANOVAs of the differences between the three conditions in mean change in liking and hunger from before and after devaluation were significantly different between groups, $p < 0.05$ (Table 5). This difference was as expected, because devaluation (i.e. changing how much the participant likes the reward) was the main purpose of the video in the paired condition. Also as expected, the change in

hunger from before devaluation to after extinction was not significantly different between groups ($p = 0.133$; Table 5). This offers support against the idea that either the paired or worms groups reduced responding by causing disgust and reduced hunger, instead supporting the idea that it was devaluation, either of the reward or of the context, respectively, that lead to decreased responding post-devaluation.

Further support for the devaluation effect caused by the video can be seen by considering the aforementioned decreases in liking after devaluation in tandem with other observations about the relationship between liking and response rate. Leaving the effects of the worms-only condition to the side for a moment, there were also significant effects of the paired condition compared to the empty bowl unpaired condition on multiple measures. The change in liking, hunger, and response rate all saw significant differences between these two conditions in independent samples t-tests ($p < 0.05$; Table 6). These differences support the hypothesis that the decreased responding post-devaluation is related to a reduction in liking, and that liking is a proxy for response rate. The idea that responding is a function of liking is central to instrumental responding (Skinner, 1938; Balleine & Dickinson, 1998a). This concept is further supported by the finding that liking before session 3 was significantly correlated with response rate in the third session ($r = 0.215$, $p < 0.05$ Table 7). Although liking immediately following devaluation and response rate during the start of extinction were not significantly correlated at the 1-tailed 0.05 alpha level for any of the groups individually ($r = 0.288$, $r = 0.21$, $r = 0.109$, Table 8), they were significantly correlated across groups ($r = 0.227$, $p < 0.05$; Table 8), further supporting liking as a proxy for response rate, and thus a predictor of the devaluation effect cause by the video.

Though the behavioral data from the behavioral paradigm offered support for the devaluation effect of the video, the behavioral data obtained from the few subjects enrolled in the fMRI paradigm did not support the notion that there is a general overreliance on the habit-based system consistent with the findings of Sjoerds in alcoholics (2013). Both of the Independent-Samples-Mann-Whitney U tests measures revealed no significant differences between smokers ($N=3$) and nonsmokers ($N=7$) (Table 9).

Similarly, the correlations performed with the scores from the personality questionnaires showed that only one personality measure, conscientiousness, demonstrated a correlation with the change in response rate before and after devaluation ($r = -0.644$) (Table 10). With such a small number of subjects, it would be premature to draw any strong conclusions. One possible explanation is individual difference in sensitivity to the type of devaluation, as one nonsmoker specifically commented after the MRI that seeing worms in food did not affect her, while one smoker indicated that the worms were “pretty gross.” However, the most powerful and most likely explanation is the small sample size. With only 10 MRI participants and only 3 smokers, it is difficult to show significant behavioral results when there is so much variation possible between participants.

The perfect negative Spearman correlation observed between the effect size of the vmPFC activity and the change in responding immediately before and after devaluation is a promising finding. The direction and strength of this correlation shows that there is greater activation in the region implicated in goal-directed behavior (the vmPFC) when behavior more closely resembles goal-directed action. However, similar to the issue with the small number of subjects contributing data for the behavioral analyses, this finding comes from such a small sample size ($N=3$) that no strong conclusions can be drawn at this time.

Despite challenges with using human subjects to study instrumental learning, it is important to find tasks that can be used to study instrumental learning in humans in order to apply the vast body of animal model research to relevant human applications. Because of its simplicity and efficacy in animal models, it is likely that the task in this study is not cognitively taxing, unlike the task used by Sjoerds and colleagues in 2013; additionally, it has now demonstrated the ability to affect instrumental responding with the content of the devaluation video and not with another factor. The demonstration of a devaluation effect also offers preliminary support for the use of this task in a 1-day paradigm in place of the 3-day MRI paradigm, in order to cut down on time and scheduling limitations of a 3-day study.

However, despite the potential efficacy of this 1-day paradigm, we will continue to use the 3-day MRI paradigm in further research. Due to the large number of unmotivated subjects that needed to be excluded, a 1-day paradigm would cause additional expensive MRIs to be performed on subjects whose data would end up having to be excluded anyway; so, by continuing to use the 3-day paradigm, we are able to filter out these low-motivation subjects.

Before the 3-day paradigm can be comfortably converted to a 1-day paradigm, a couple of improvements can be made with the present task. One potential improvement that could address the lack of a significant difference between groups would be to train participants on the task for much longer. In their similar paradigm, Tricomi and colleagues (2009) had their undertrained group perform two 8-minute training sessions on one day for 16 minutes of training, while their overtrained group performed four 8-minute training sessions each day for 3 days for a total of 96 minutes of training. Tricomi and colleagues (2009) saw a clear devaluation effect in the 1-day group, with responding persisting into extinction in the 3-day group, showing clear goal-directed and habit-based behavior, respectively. To compare, our subjects each had 6 minutes of training each day for a total of 18 minutes of training; this suggests our subjects would be more likely to demonstrate goal-directed behavior based on training amount alone. It might be possible to find the amount of training that shows goal-directed behavior for nonsmokers and habit-based behavior for smokers. Additionally, literature suggests that the addition of a stressor could induce habit-based behavior in those with drug addiction, which could be examined in future studies with similar or greater amounts of training (Schwabe & Wolf, 2009; Schwabe & Wolf, 2011).

It is also important to note that responding never fully extinguished in the unreinforced trials following devaluation. One possible solution is suggested by the finding of a similar task paradigm used in Dr. Eric Thraillkill's currently running behavioral study (data not shown). In this paradigm, the vending machine stops tilting during extinction, and the behavior does extinguish. While fully

extinguishing is not necessary to see a devaluation effect, it would give stronger support to any findings obtained, and would hopefully, when combined with greater overtraining on the task, more clearly demonstrate those more strongly dependent upon habit-control, be it those with drug addiction, stressed individuals, or those with other factors contributing to great reliance on habit-control.

- Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *34*(2), 77-98.
- Balleine, B. W., & Dickinson, A. (1998a) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*, 407-419.
- Balleine, B. W., & Dickinson, A. (1998b). The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. *Animal Learning Behavior*, *26*, 46–59.
- Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology Reviews*, *35*, 48-69.
- Balleine, B. W., Morris, R. W., & Leung, B. K. (2014). Thalamocortical integration of instrumental learning and performance and their disintegration in addiction. *Brain Research*, *1628*, 104-116.
- Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation*, *10*, 445-466.
- Bouton, M. E., Todd, T. P., Vurbic, D., & Winterbauer, N. E. (2011). Renewal after the extinction of free operant behavior! ! *Learning & Behavior*, *39*, 57-67.
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, *53*(1), 1-15.
- Liljeholm, M., Tricomi, E., O'Doherty, J. P., & Balleine, B. W. (2011). Neural correlates of instrumental contingency learning: Differential effects of action–reward conjunction and disjunction. *The Journal of Neuroscience*, *31*(7), 2474-2480.
- Lovibond, P. F. (1981). Appetitive pavlovian-instrumental interactions: Effects of inter-stimulus interval and baseline reinforcement conditions. *The Quarterly Journal of Experimental Psychology Section B*, *33*(4), 257-269.

- McNamee, D., Liljeholm, M., Zika, O., & O'Doherty, J. P. (2015). Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate fMRI study. *The Journal of Neuroscience*, *35*(9), 3764-3771.
- Morris, R. W., Quail, S., Griffiths, K. R., Green, M. J., & Balleine, B. W. (2015). Corticostriatal Control of Goal-Directed Action Is Impaired in Schizophrenia, *Biological Psychiatry*, *77*(2), 187-195.
- Mintzer, M. Z., & Stitzer, M. L. (2002). Cognitive impairment in methadone maintenance patients. *Drug and Alcohol Dependence*, *67*(1), 41-51.
- Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, *29*(22), 7191-7198.
- Schwabe, L., & Wolf, O. T. (2011). Stress-induced modulation of instrumental behavior: From goal-directed to habitual control of action. *Behavioural Brain Research*, *219*(2), 321-328.
- Scott, J. C., Woods, S. P., Matt, G. E., Meyer, R. A., Heaton, R. K., Atkinson, J. H., & Grant, I. (2007). Neurocognitive effects of methamphetamine: A critical review and meta-analysis. *Neuropsychology Review*, *17*, 275-297.
- Siegel, S. (1999). Drug anticipation and drug addiction. The 1998 H. David Archibald Lecture. *Addiction*, *94*(8), 1113-1124.
- Smith, S. M., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23*(1), S208-219.
- Sjoerds, Z., de Wit, S., van den Brink, W., Robbins, T. W., Beekman, A. T., Penninx, B. W., & Veltman, D. J. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Translational Psychiatry* *3*(337), 1-8.
- Skinner, B. F. (1938). *The Behavior of organisms: An experimental analysis*. New York: Appleton-Century.

- Skinner, B. F. (1979). *The shaping of a behaviorist: Part two of an autobiography*. New York: Knopf.
- Thrailkill, E. A., & Bouton, M. E. (2014). Contextual control of instrumental actions and habits. *Journal of Experimental Psychology: Animal Learning and Cognition*, *41*(1), 69-80.
- Tournier, J. D., et al. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage*, *23*(3), 1176-1185.
- Tournier, J. D., Calamante, F., & Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage*, *35*(4), 1459-1472.
- Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *The European Journal of Neuroscience*, *29*(11), 2225-2232.
- Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *The Journal of Neuroscience*, *27*(15), 4019-4026.
- Verdejo-García, A., Bechara, A., Recknor, E. C., & Pérez-García, M. (2006). Executive dysfunction in substance dependent individuals during drug use and abstinence: An examination of the behavioral, cognitive and emotional correlates of addiction. *Journal of the International Neuropsychological Society*, *12*(3), 405-415.