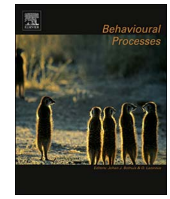
Behavioural Processes 149 (2018) 16–26

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc



The temporal dynamics of waiting when reward is increasing

Michael E. Young*, Lisa Vangsness, Anthony W. McCoy

Kansas State University, USA

ARTICLE INFO

Keywords:
Impulsivity
Interresponse times
Waiting
Video game

ABSTRACT

The temporal dynamics of waiting are complex. The present study used a video game involving contingencies that produced differential reinforcement of wait times by arranging for the magnitude of the reward to be related to the duration of each inter-response time. In previous research, when outcomes were gradually increasing in value from a minimum to a maximum, two modes of behavior are observed: waiting as little as possible before cashing in (i.e., responding rapidly) or waiting until the maximum is available (i.e., responding slowly). When outcomes were either a smaller sooner or larger later reward, two modes of behavior were again observed which corresponded to choosing either reward immediately after it was available. In the present study, outcome values increased linearly for a period of time, leveled off, increased abruptly, and then increased linearly. This configuration produced three modes of wait times in people – responding immediately, responding immediately after the abrupt increase, and responding when the maximum possible reward was achieved on each trial. Three factors were evaluated as possible causes of each behavioral mode, a desire for immediacy or action, the trade-off between molecular and molar maximization, and ease of responding.

1. Introduction

In every situation, people must wait for outcomes. The delays to an outcome may range from short imperceptible delays, like that between pressing a key on a keyboard and the appearance of the corresponding character on the screen, to delays involving days or months like that between placing an online order and receiving the purchased product. While waiting, there may be an external indicator of progress toward delivery of the outcome (e.g., a continuous progress bar on a computer or discrete email updates that your order was received or your product has shipped) or not. Furthermore, in some circumstances the outcome may only be available at the end of the waiting period (e.g., the arrival of a purchased product), or lesser or lower quality outcomes may be available earlier as for a savings account, an internet search, or a ripening banana.

When waiting for the delivery of a desirable outcome, one has the option to wait or not. This involves the tradeoff between some smaller sooner outcome and a larger later outcome that is provided at the end of the wait period. When studying this tradeoff, researchers typically note the probability that a subject will take the smaller sooner reward rather than the larger later reward (e.g., Johnson and Bickel, 2002; Odum, 2011; Rachlin, 2006). When the specific wait times are examined, a subject typically takes the chosen reward immediately after it is available (Young and McCoy, 2015); there is little reason to wait longer than one has to after the desired reward is presented.

The situation becomes more complex, however, when the value of the reward is not constant during the wait period. How long should an investor wait before withdrawing funds from a savings account that is continuously increasing in value? How long should a soldier wait for a better shot? And, how long should someone forage for information or food before moving on to another patch or to other tasks? In each of these cases, the answer depends not only on an organism's evolving needs, but also on the dynamics of the changes in value. For example, is the outcome value increasing linearly, negatively accelerating thus producing diminishing returns over time, or positively accelerating so that waiting is increasingly more beneficial? Each of these patterns of increase should prompt different patterns of waiting.

Predicting behavioral dynamics in the presence of within-trial reward dynamics is particularly challenging because conventional studies of choice are based on the assumption that each outcome has a relatively constant but unknown value that must be experienced or sampled to learn this value (e.g., Mazur, 1992; Rescorla and Wagner, 1972; Sutton and Barto, 1981, 1991, 1998). When an outcome's value changes within a trial, however, optimal behavior requires an organism to model these value dynamics and to adapt its behavior to maximize the molar reward rate (i.e., the long-term expectation of the number of rewards obtained per unit of time). Theories of this process will likely have more in common with theories of motor learning than traditional reinforcement learning (Pearson and Platt, 2013). The lack of behavioral data involving within-trial stimulus and value dynamics makes it

* Corresponding author.
E-mail address: michaelyoung@ksu.edu (M.E. Young).

https://doi.org/10.1016/j.beproc.2018.01.015
Received 6 July 2017; Received in revised form 27 September 2017; Accepted 19 January 2018
Available online 31 January 2018
0376-6357/

premature to propose such a theory, but reinforcement models assume that an organism adapts its behavior to optimize the molar reward rate (Sutton and Barto, 1998). Given that people routinely fail to wait when it is optimal to do so (Logue, 1988; Mazur and Logue, 1978; Mischel et al., 1989), there must be other aspects of the environment that produce these apparently maladaptive behaviors. In other words, what other aspects of the environment are sufficiently rewarding that they will shift wait times away from optimal behavior? To study the effect of within-trial stimulus and reward dynamics on waiting, our laboratory prepared an environment where participants would be sufficiently engaged to make many of these decisions in a short period of time.

In the following sections, we will begin by considering the factors that may change how long a subject will wait when the value of an outcome is changing within trial. This work will then be examined within the larger context of schedules of reinforcement that differentially reward particular wait times (e.g., differential reinforcement of low rates of responding). We then present data showing multiple wait time modes in our earlier work and discuss the possible role of effort in driving the occurrence of one of these modes. Finally, we consider another factor that could produce one of these modes – a discontinuity in the growth of a commodity when its value levels off. These alternative explanations will drive the design of our experiment.

2. Waiting when rewards are rapidly changing

In the predominant task used in our laboratory, participants play a first-person-shooter video game (Young and McCoy, 2015; Young et al., 2011; Young et al., 2013b). While playing, each participant sees a "charge" bar in the middle of the screen that gradually fills over a period of 10 s. The player can either wait a short period of time before firing their weapon and thus do less damage because the bar is not full, or the player can wait longer before firing in order to do more damage. Two factors complicate their decisions. First, the player can immediately take another shot such that the behavior that maximizes the rate of reward is not always the one that maximizes the reward for the present shot. For example, if waiting 2 s produces 20 points of damage but waiting the full 10 s produces 40 points of damage, then the player should take 5 shots every 2 s over the 10 s period rather than only 1 shot because the total damage would be 100 points rather than 40 points. Second, the way in which the bar fills will change on occasion such that the wait time that maximizes the reward rate changes.

To better understand the factors driving waiting, the primary dependent variable in our studies is the amount of time that people are willing to wait between shots – the interresponse time (IRT). A consistent challenge we have faced in assessing individual differences and the effects of various environmental manipulations on IRTs is the complexity of this variable's distribution. In our early research (Young et al., 2011; Young et al., 2013a,b), IRTs were highly bimodal with short IRTs produced by responding rapidly (usually by holding down the response key to avoid muscle fatigue) and long IRTs produced by waiting for the maximum possible outcome (a full charge bar) which was usually available after 10 s. This distribution was not surprising given that an optimality analysis revealed that our changing task contingencies favored responding as rapidly as possible in some conditions and waiting the full 10 s in other conditions. However, these two modes were commonly observed even under conditions when only one or the other was explicitly reinforced (cf. Arbuckle and Lattal, 1988). This observation suggested that factors in addition to optimizing molar reward rate were determining wait times. To shed light on the issue, we first turned to the previous literature on IRTs (Blough, 1963; Ferster and Skinner, 1957; Shimp, 1969).

The wait times in our task are functionally equivalent to an IRT because each opportunity to respond is immediately followed by another. In our gaming task, each IRT produces a different level of reward as a function of the duration of the IRT – longer IRTs produce a larger immediate reward than shorter IRTs. Thus, our task designs are similar to those used in Skinner's studies of differential reinforcement of response rates (Ferster and Skinner, 1957) in which reinforcement is provided only if the IRT is above (differential reinforcement of low rates), below (differential reinforcement of high rates), or within a specified range (differential reinforcement of pacing). In Skinner's work on rate-based schedules, reward was discrete and provided when a particular range of IRTs or series of IRTs had been produced. Our video game preparation has some unique characteristics, however, because there is a continuous range of reward levels that are contingently related to the duration of each IRT, and there is an external cue (the charge bar) that indicates the current reward level.

Although we can precisely control the experimentally-determined IRT contingency, there are many contingencies in an environment that are not intentionally manipulated (Lattal, 1995). Skinner's (1948) observation of superstitious behavior in the pigeon represents an early demonstration of unplanned contingencies that emerge and alter behavior. Furthermore, many types of rewards exist in any one environment, each one of which may reinforce a different behavior. While playing a first-person-shooter video game, a player simultaneously may desire to destroy an opponent, obtain a high score, complete a game challenge quickly, and minimize the risk of their character dying. It is unlikely that the same behavior will be consistent with achieving all of these goals simultaneously.

Thus, it is likely that multiple factors are affecting participant IRTs in our task. We hypothesized that three factors contribute to multiple modes in the distribution of IRTs. First, the distributions should be determined by the experimenter-manipulated task contingencies and thus optimize molar reward rate (Rachlin et al., 1981). This factor has been present in each of our published studies and clearly documented by the observed shifts in IRTs as the task contingencies changed. A second factor that we have suspected since the beginning of this line of research is the relative ease of responding. To avoid response fatigue, we have allowed participants to automatically generate very rapid responses (every 0.25 s) by holding down the response key. The ease of generating fast responses coupled with the satisfaction of action is likely to produce an overall preference for short IRTs regardless of the circumstances. A third factor revealed in the present study for the first time is the desire to achieve the maximum possible outcome on a given trial even if waiting for that outcome provides no molar benefit (and perhaps even when it is detrimental in the molar sense, Shimp, 1966). The desire to achieve the maximum molecular reward can have opportunity costs because of foregone rewards that could have been obtained during waiting. More concretely, if an outcome is gradually increasing in value toward its maximum but at too slow of a rate to reward additional waiting, then the subject should not wait but instead accept a smaller reward earlier. But, if a desire to achieve the maximum molecular reward is also driving behavior, then we may observe a third IRT mode at the time that this maximum is achieved.

In essence, we propose that three classes of behavior occur due to three different rewards being present in our task: maximizing molar reward rate (maximizing the rate of destroying targets in our gaming preparation), engaging in low effort (or, perhaps, the reward associated with simply acting), and maximizing molecular reward amount (obtaining the highest amount of weapon charge possible without consideration of the molar effects). Because in our earlier studies the optimal IRTs were either to respond rapidly or to wait for the maximum amount, these three factors were inherently confounded. For the purposes of our research questions, this confounding was not a problem. Our interest was in individual differences in the propensity to wait and the environmental factors that affect that propensity in all individuals. Whether the variance in IRTs across people and situations was due to greater optimality, desire for immediacy, or desire for maximum molecular outcomes was not of central concern. Except in the presence of ceiling (always waiting) or floor (always responding rapidly) effects, we could distinguish between people's optimality and an overall tendency toward waiting that allowed us to address our central research