Decision-tree analysis of control strategies

Article *in* Psychonomic Bulletin & Review · October 2014

DOI: 10.3758/s13423-014-0732-0 · Source: PubMed

Romann Weber
Disney Research

DECISION-TREE ANALYSIS OF CONTROL STRATEGIES

Romann M. Weber

California Institute of Technology

Brett R. Fajen

Rensselaer Polytechnic Institute

Author Note

Work on this project was supported by a Rensselaer Humanities, Arts and Social Sciences Graduate Fellowship awarded to the first author. The first author would like to thank his coauthor and Wayne Gray, Mei Si, and Mark Changizi for their guidance during the development of this work. Both authors would like to thank this journal's editorial staff and an anonymous reviewer for their helpful comments. Correspondence concerning this article should be addressed to Romann M. Weber, MC 228-77, California Institute of Technology, Pasadena, CA 91125. email: rweber@caltech.edu

Abstract

A major focus of research on visually guided action is the identification of control strategies that map optical information to actions. The traditional approach has been to test the behavioral predictions of a few hypothesized strategies against subject behavior in environments in which various manipulations of available information have been made. While important and compelling results have been achieved with these methods, they are potentially limited by small sets of hypotheses and the methods used to test them. In this study, we introduce a novel application of data-mining techniques in an analysis of experimental data that is able to both describe and model human behavior. This method permits the rapid testing of a wide range of possible control strategies using arbitrarily complex combinations of optical variables. Through the use of decision-tree techniques, subject data can be transformed into an easily interpretable, algorithmic form. This output can then be immediately incorporated into a working model of subject behavior. We tested the effectiveness of this method in identifying the optical information used by human subjects in a collision-avoidance task. Our results comport with published research on collision-avoidance control strategies while also providing additional insight not possible with traditional methods. Further, the modeling component of our method produces behavior that closely resembles that of the subjects upon whose data the models were based. Taken together, the findings demonstrate that data-mining techniques provide powerful new tools for analyzing human data and building models that can be applied to a wide range of perception-action tasks, even outside the visual-control setting we describe.

DECISION-TREE ANALYSIS OF CONTROL STRATEGIES

Introduction

An animal's perception of the environment and ability to control action is made possible by the availability of information contained in ambient energy arrays (Gibson, 1986). An understanding of the components within these energy arrays that make them informative often yields useful insights into how animals perceive and act in a particular context. In the study of *visually guided action*, this often entails an analysis of optic flow fields, with the goal of identifying optical variables that specify action-relevant properties and prescribe how to act to bring about a particular goal. This undertaking has led to the formulation of numerous control laws for steering, fly ball catching, braking to avoid a collision, and a variety of other tasks (see Fajen, 2005b, Warren, 1998 for reviews).

Discovering new sources of information through the analysis of optic flow fields alone is extremely challenging. The aim of this paper is to introduce a novel approach to identifying sources of optical information for visual control. This approach makes use of a technique from data mining called decision-tree learning and differs from previous approaches in that behavioral data are used at an earlier stage of the process. Rather than relying entirely on the researcher's intuitions to identify candidate informational variables, the method we describe lets the behavioral data tell the researcher what those informational variables may be. As we show, this can lead to new insights into the information-based control of action. Furthermore, although we demonstrate the method within the context of visual control, it is hardly limited to this domain. In principle, essentially any experimental variable from multiple domains of interest can be investigated in the manner we describe here.

Information and Laws of Control: A Brief Review

Control laws (or control strategies) often take the form of mappings from information variables that specify action-relevant states of the environment to action variables that are controlled by the actor (Warren, 1988, 1998). The information variables in many control law models are *optical invariants* in the sense that they invariantly specify action-relevant states across variations in commonly encountered conditions. Among the most well known applications of one of these information variables is Lee's (1976) *tau-dot* ($\dot{\tau}$) model of braking, where $\dot{\tau}$ is the first time derivative of τ, the latter defined as the ratio of the optical angle (θ) subtended by the approached object to its rate of expansion ($\dot{\theta}$). By regulating deceleration to keep $\dot{\tau}$ as close as possible to the critical value of -0.5, the observer will come to a stop at the object (Yilmaz & Warren, 1995). This works because $\dot{\tau}$ invariantly specifies the sufficiency of the observer's current rate of deceleration. That is, regardless of the size of the approached object or the speed of approach, the value of $\dot{\tau}$ relative to -0.5 specifies whether the current rate of deceleration is insufficient ($\dot{\tau} <$ -0.5) or excessive ($\dot{\tau} >$ -0.5).

Control law models of other tasks follow a similar logic. For example, in Chapman's (1968) *optical acceleration cancellation* (OAC) model of fly ball catching, the acceleration of the tangent of the ball's elevation angle invariantly specifies the sufficiency of the outfielder's current running speed for arriving at the landing location in time to catch the ball. In the *constant bearing angle* model of interception (Chapman, 1968; Fajen & Warren, 2004; Lenoir, Musch, Thiery, & Savesbergh, 2002), the rate of change of the bearing angle of the moving target invariantly specifies the sufficiency of the observer's current locomotor speed (or direction) for intercepting the moving target.[1]

The structure of control laws in perception and action is intuitively similar to the concept of control in an engineering context, with some reference quantity either needing to be maintained around a target value or prescribing a change of behavior when a threshold value is reached, much like a car's cruise control or an air conditioner's thermostat. Unlike

[1]In other models of visually guided action, information variables invariantly specify *affordances*, such as whether it is within one's capabilities to safely avoid a collision with an object in the path of motion by braking (Fajen, 2005a, 2005c), or whether it is within one's capabilities to intercept a moving target on foot or pass between a pair of moving obstacles (Fajen, 2013).

control engineers, behavioral control-law researchers have the additional challenge of not knowing ahead of time the reference variable or variables to be controlled. For most behaviors of interest, the choices are hardly obvious.

Current Approaches to Finding Control Laws

When researchers seek to determine the control strategy for a particular task, they usually begin by listing the possibilities. This often involves an analysis of the geometry and physics of the task at hand, assuming that the task is sufficiently simple to permit such an analysis. Candidate control strategies are then tested by analyzing performance on the given task under various conditions. Experiments are often conducted in virtual or simulated environments, allowing researchers to more easily manipulate sources of information available to subjects and monitor how performance is affected by those manipulations.

We will illustrate this standard approach using a study by Fajen and Devaney (2006), in which subjects performed an "emergency braking" task in a virtual environment. Subjects were instructed to initiate braking at just the right moment so that they would come to a stop directly in front of a target, a row of stop signs. The brake in the experiment differed from a standard brake in that it was either off, in which case deceleration was zero, or on, in which case the rate of deceleration was 10 m/s^2. As such, the task required subjects to initiate braking at precisely the right moment, namely when the fixed rate of deceleration needed to stop at the target was equal to the fixed rate of deceleration provided by the brake (10 m/s^2).

In terms of spatial variables, the fixed rate of deceleration needed to stop at the target (which Fajen and Devaney (2006) referred to as *ideal deceleration* or d_{ideal}) is equal to $v^2/2z$, where v is the current speed and z is the current distance from the intended stopping location. Thus, the task requires one to initiate deceleration precisely when $v^2/2z$ is equal to the fixed rate of deceleration of the brake. This condition can also be expressed in terms of optical variables by relying on the fact that $\tau = \theta/\dot{\theta} \approx z/v$ and the fact that the *global optic flow rate* (GOFR) is directly proportional to v:

$$d_{\text{ideal}} = \frac{v^2}{2z} \propto \text{GOFR}\frac{\dot{\theta}}{\theta} = \frac{\text{GOFR}}{\tau} \tag{1}$$

Fajen and Devaney (2006) considered three possible braking strategies and derived performance consequences of using each one. Specifically, the authors showed that if subjects use a strategy of applying the brake when either $\dot{\theta}$ or $\dot{\theta}/\theta$ reaches some constant threshold value, then one would expect to see a tendency to apply the brake too early or too late under certain conditions, while using $\text{GOFR}\frac{\dot{\theta}}{\theta}$ would not introduce a timing bias. These patterns of brake timing follow predictable functional forms when plotted versus the size (specifically, the radius) of the approached object and versus the initial speed of approach. Subjects' actual braking behavior can then be plotted against these predicted curves in an effort to distinguish one strategy from another (see Figure 1).

As shown in Figure 1A (which corresponds to Experiment 1A, in which sign radius was varied and initial speed was fixed), there is strong evidence to suggest that subjects initially relied on the expansion rate, $\dot{\theta}$, alone, as revealed by the bias in their braking behavior when the sign radius was altered. This behavior changed, however, by the end of the experiment, and the resulting braking bias appears in between what is suggested by the "pure" strategies. The other plots in Figure 1 show predictions and performance under other conditions that were tested by Fajen and Devaney: when sign radius was varied and the textured ground plane (i.e., the source of GOFR information) was absent (1C), when initial speed was varied with the ground plane present (1B), and when initial speed was varied with the ground plane absent (1D). These plots show similar effects, with performance in early trials aligning fairly well with one of the theoretical curves before drifting to an "in-between" strategy in later trials.[2]

[2] One might wonder whether the observed changes in behavior over blocks reflect calibration to the strength of the brake rather than attunement to different optical variables. This is unlikely for the following reason: In a different study using a similar task, Fajen (2007) found that subjects were able to rapidly

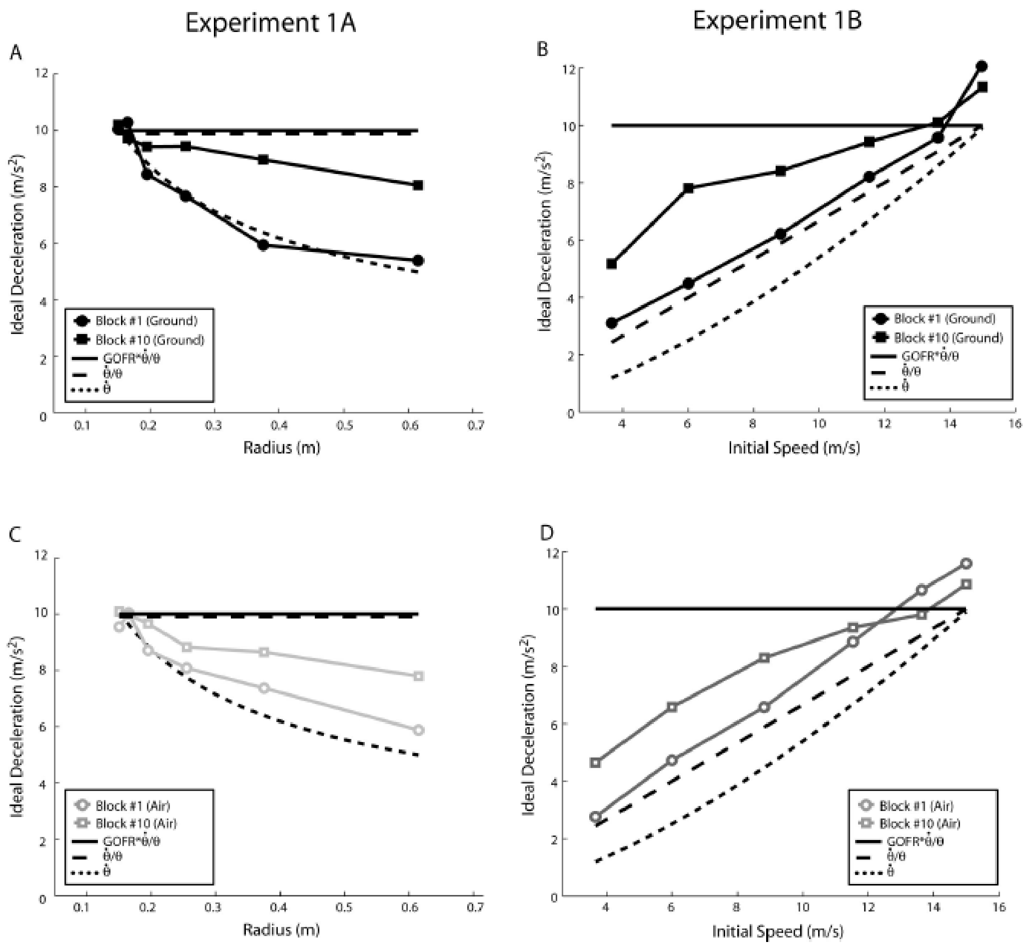


Figure 1. Mean ideal deceleration at brake onset as a function of target radius (A and C) and initial speed (B and D). Trials with the ground plane present are shown at top, while those without (the "Air" condition) are at bottom. Figure 6 from Fajen and Devaney (2006).

Although Fajen and Devaney (2006) focus on the braking task, their study is representative of the most common approach to the analysis of control laws in visually guided action. This approach can be summarized as follows:

1. Generate a set of candidate control strategies, motivated by previous research or first principles.

2. Show that each strategy or class of strategies will produce behavior that, under certain conditions, is distinct from the behavior produced using other strategies.

3. Propose a quantitative measure of this behavior or performance.

4. Gather experimental data under the various conditions considered.

5. Determine whether performance differs across conditions in a manner that would be expected if subjects were using one of the candidate control strategies.

Limitations of the Standard Approach

The standard approach, while powerful in some respects, also has some limitations. First, formulating candidate control strategies from first principles is quite challenging. Although analyses of the geometry and physics of the task sometimes lead to the formulation of control strategies (e.g., Lee's $\dot{\tau}$ strategy, Chapman's OAC strategy), it has proven difficult to identify informational variables and control strategies for other, more complex tasks. Even in cases in which candidate control strategies can be generated, it is hardly obvious how to create experimental manipulations that pit one strategy against another.

Second, numerous studies have demonstrated that observers often rely on non-invariants and that the particular optical variables upon which they rely to guide

recalibrate to sudden increases or decreases in the strength of the brake. Recalibration was mostly complete within about ten to fifteen trials. Because the conditions used in Fajen (2007) were so similar to those used in Fajen and Devaney (2006), it is reasonable to assume that the process of learning the strength of the brake was complete by the end of Block 1. Therefore, the changes in Figure 1 likely reflect attunement to more effective optical variables rather than calibration.

action can change—as a consequence of practice, as a function of the range of conditions that are encountered, and as a function of the dynamics of the controlled system (Fajen, 2008a, 2008b; Fajen & Devaney, 2006; W. Li, Saunders, & Li, 2009; Smith, Flach, Dittman, & Stanard, 2001; Stanard, Flach, Smith, & Warren, 2012). This process is sometimes referred to *perceptual attunement*.[3]

Researchers can include non-invariants in their set of candidate optical variables, as Fajen and Devaney (2006) did in the aforementioned example. When human behavior is aligned with the predictions of one of these non-invariants, that provides evidence, albeit not conclusive evidence, for the use of that variable. However, it is not uncommon for the human data to fall in between the predictions of two of the candidate variables. The data from Block 10 in Fajen and Devaney (2006), shown in Figure 1, is one such example. Such findings can be difficult to interpret because it is not clear what optical variable subjects are using.

This becomes especially problematic when it comes to studying perceptual attunement as a process in which observers gradually transition from relying on one optical variable to another. If the data suggest that an observer was attuned to $\dot{\theta}$ early in practice and $\dot{\theta}/\theta$ later in practice, what can be said about the transition from relying on $\dot{\theta}$ to relying on $\dot{\theta}/\theta$?

Smith et al. (2001) proposed that the gradual changes that accompany perceptual attunement could be understood in terms of changes in a linear margin in a state space of optical variables. In their study, subjects timed the release of a pendulum to hit an approaching ball, and the dimensions of their state space were the optical angle (θ) and expansion rate ($\dot{\theta}$) of the approaching ball. By representing the linear margin using the equation $\dot{\theta} = a + b\theta$, Smith et al. could capture a continuum of strategies by varying the parameters a and b. For example, setting b to zero and a to some value greater than zero

[3] This has been referred to elsewhere in the literature as the *education of attention*; see Jacobs and Michaels (2007).

corresponds to a $\dot{\theta}$ strategy (i.e., release the pendulum when $\dot{\theta}$ reaches a); setting a to zero and b to some value greater than zero corresponds to a $\dot{\theta}/\theta$ strategy (i.e., release the pendulum when $\dot{\theta}/\theta$ reaches b). Thus, the transition from $\dot{\theta}$ to $\dot{\theta}/\theta$ can be understood as a gradual change that involves the tuning of parameters a and b.[4]

Because Smith et al. (2001) restricted their analysis to linear margins in optical state space, the possibility that subjects use a source of information captured by some non-linear combination of variables cannot be considered. In principle, one could generalize this approach to allow for non-linear combinations. However, it is unclear how to choose the forms of the functions that would be considered.

It is with these limitations in mind that we considered the need for a new approach to the analysis and modeling of visual control strategies. The approach involves mining models of visually guided action directly from human data and is a specific application of a broader data-mining theme explored in more detail by Weber (2013).

Mining Models from Data

Data mining is being increasingly used as a means of dealing with intrinsically high-dimensional data sets, especially when the information that has been collected is in a database built without specific tests or hypotheses in mind. As a result, data-mining methods make few assumptions about the data they process, which stands in contrast to the various assumptions that are necessary for the application of the most commonly encountered statistical tests in the experimental psychology literature.

Data mining has become especially well known for its applications in commercial settings (Provost & Fawcett, 2013). With an increasing amount of information about transactions and browsing behavior being collected electronically, companies look for patterns in this information that will help target advertising or product recommendations

[4] An interesting, related take on this issue comes from Jacobs, Michaels, and colleagues (see Jacobs and Michaels (2007) for an introduction), who consider the attunement process as a path through an *information space*. We consider their approach in the discussion at the end of the paper.

to users.

The possibilities of data mining are just beginning to become clear to researchers within experimental psychology. Recent indicators of the field's nascent interest in these methods include the symposium dedicated to the topic at the 2012 Annual Meeting of the Psychonomic Society (Psychonomic Society, 2012),[5] the 2014 course in exploratory data mining (the latest of several) sponsored by the American Psychological Association (American Psychological Association, 2014), and the advocacy of an expanding toolkit to practitioners of experimental psycholinguistics (Baayen & Cutler, 2005).

Mining Control Laws from Behavioral Data

How can data mining be used to discover control laws for visually guided actions? When a subject in an experiment executes an action on the basis of optical information, he or she is effectively *labeling* that information. This "label"—a record of the subject's behavior—then becomes the target for the classification rules that data-mining algorithms are built to learn.

To illustrate this point, recall the braking example introduced earlier. In this task, subjects move along a straight path toward a target object, and the goal is to come to a stop as close as possible to the object without hitting it. Let us consider a simplified version of the task in which there is only one action to take, namely the execution of a single braking maneuver at maximum brake force precisely timed to come to a stop just before reaching the object. In this case, the data being recorded include the time at which the brake was initiated (the behavior we are interested in modeling) along with the values of the environmental variables at that time (the *predictors*). However, if we are to use a classification algorithm to detect the signature of a control strategy, we need to make sure that the data fall into at least *two* labeled categories. This requires a slight adjustment to

[5]Even so, the applications advocated in that symposium and elsewhere were largely in line with traditional approaches to data mining and focus on research into text corpora, image statistics, and certain aspects of behavioral economics.

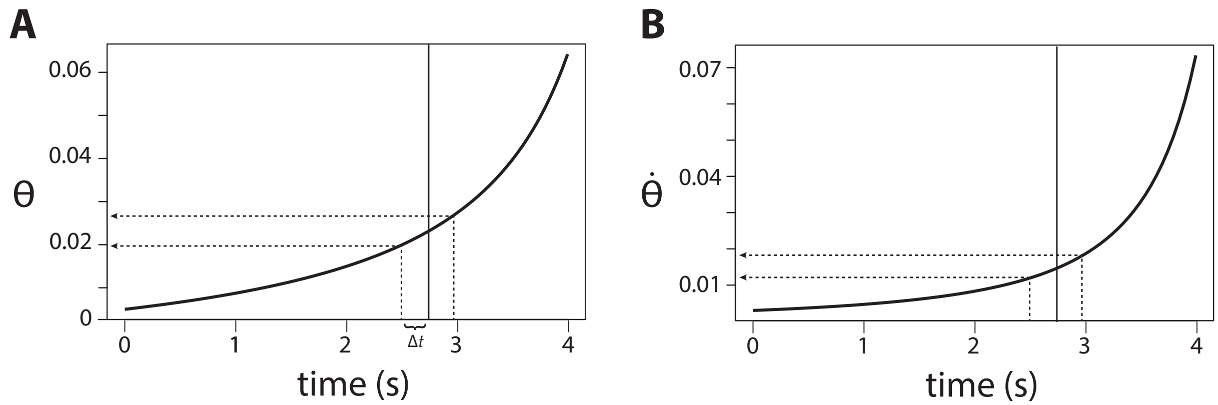


Figure 2. Labeling data. The thick black curves correspond to the values of θ (A) and $\dot{\theta}$ (B) over time during constant-velocity approach to a target. The solid vertical lines indicate the point of brake application. The dashed lines indicate the pre- and post-braking data points, which are time-shifted relative to the recorded braking point.

the raw data we have collected, which we will now explain.

There are many possible measures of the optical scene, but for simplicity we will consider just two for the moment: the optical angle of the target object, θ, and its rate of expansion, $\dot{\theta}$. The black curves in Figure 2 show how θ and $\dot{\theta}$ change over time as the subject approaches the target on a sample trial.[6] The vertical line indicates the moment at which the subject hit the brake.

The assumption is that braking was initiated at that moment because a threshold value of some optical variable, variables, or function thereof was reached. We can then regard the paired values of θ and $\dot{\theta}$ prior to braking as falling in a "do not brake" zone, since these values did not trigger the application of the brake.[7] Presumably, the paired values of θ and $\dot{\theta}$ that would have been registered after this point *had the brake not been*

[6]These quantities are easily computed from spatial data—i.e. distances and velocities—using simple trigonometry.

[7]This necessarily ignores the lag between the registering of a threshold and the action it triggers. However, since a behavioral strategy should account for this lag and incorporate it into decision making, it is reasonable to regard the recorded point as occurring at the "effective" threshold.

applied would fall within a "brake" zone. If we look forward and backward in time on our plots, shifting left and right by some small value Δt, we can choose values of θ and $\dot{\theta}$ that represent these two zones.

Said in more detail, consider the time series of θ and $\dot{\theta}$ for the case in which no braking action takes place—that is, as if the subject drives right into the target without ever slowing. Clearly, a subject who actually applies the brake will produce a $(\theta, \dot{\theta})$ series that differs from the hypothetical "inaction" series only *after* the time t at which the brake is applied. The timing of the brake application is up to the subject, and it is this timing decision alone that creates any difference between the subject's data and the "inaction" data. For whatever our choice of Δt (we chose 90 ms), we can wind the clock back from t and record the values of $\theta(t - \Delta t)$ and $\dot{\theta}(t - \Delta t)$ to represent the "do not brake" zone. The assumption is that whatever function describes the control law, these lagged values of θ and $\dot{\theta}$, when considered as potential arguments in that function, do not produce a result that exceeds the threshold required to initiate the braking decision. Similarly, we can look forward in time from t and pretend the braking never happened, recording $\theta(t + \Delta t)$ and $\dot{\theta}(t + \Delta t)$ from the "inaction" data to represent the "brake" zone. Here the assumption is that, since braking *did* occur at t, the $t + \Delta t$ values of these variables in the "inaction" data must lead to a result that *exceeds* the threshold required to cause the subject to initiate the brake.

In short, we try to find the threshold that triggers the decision by looking at the data from times *on each side of the decision* while discarding the data from the time of the decision itself. This is illustrated by the dashed vertical lines to the left and right of the solid vertical line in Figure 2. We record *these* points and discard the one we originally recorded. As we gather data from more trials, we create a new data set composed of points with attributes (θ and $\dot{\theta}$) that fall into one of two classes ("do not brake" or "brake"). However many decisions the subject makes, we will get twice as many data points, creating an even split between examples from the "do not brake" and "brake" zones.[8] The half of the data from before the braking time are labeled as "do not brake," and the half from after the braking time are labeled as "brake." These labeled data points make up the training set for the classifier. Next, we explain how the control laws that generated these data can be mined using a technique called *decision tree learning*.

Decision Tree Learning

Almost any algorithm or procedure can be completely described in an intuitive graphical format with a flowchart (Downing, Covington, & Covington, 2000). Decision-tree models allow one to view information-action patterns in data in much the same way. This is not to imply the use of some kind of flowchart by the perceptual or motor system. Rather, decision-tree models are a way of formally capturing control strategies in a manner that allows for visualization and is more intuitive than a mathematical equation. We will show that it is possible to gain new insight into behavior by analyzing the *structure* of these models when they are built from data drawn from different points in the experiment and in various experimental conditions. It is this deeper analysis that permits one to observe the strategic shift of perceptual attunement with task experience.

The aim of a decision tree analysis is to predict a category for a data point based on the values of its attributes.[9] Each node of the tree corresponds to a "measurement" of a particular attribute—that is, a question, often of the "yes" or "no" variety, that one could ask about that attribute's value (e.g. "Is x less than 2.5?"). The result of that measurement determines the edge followed from that node, leading to either another node—and hence another measurement—or to a *leaf*, which corresponds to the predicted category. In the example given in this work, we are trying to predict whether a subject will brake or not brake given the values of certain optical variables of interest—θ and $\dot{\theta}$, for instance.

Decision trees have many qualities to recommend them. In addition to their being

[8]Each experimental block contained 30 trails, leading to 60 data points, 30 each from each "zone."

[9]These variables are also referred to in various parts of the literature as *inputs*, *features*, or *predictors*.

easy to understand and interpret, decision trees are also relatively inexpensive computationally (Chattamvelli, 2011; Witten, Frank, & Hall, 2011). They work well on both large and small data sets and are easily "pruned" to avoid overfitting and to aid generalization to new data. Exact implementation details and pruning methods differ somewhat depending on the algorithm used, but the general principles are very similar across methods. Decision trees are also quite effective at identifying and eliminating useless attributes. This gives the user the option of creating arbitrary nonlinear functions of the most relevant attributes identified in the first round of modeling for a new round of testing. Lastly, decision trees are considered *nonlinear* classifiers, because they can partition a data set that is not linearly separable (Theodoridis & Koutroumbas, 2009).

Unlike logistic regression, a classification method quite familiar to perception researchers, decision trees do not impose *a priori* functional constraints on the variables of interest in the model determining class membership. With logistic regression, the logarithm of the *odds* of class membership is most commonly represented by a linear function of the variables of interest. This necessarily implies that as the magnitude of this function increases, the probability of class membership increases along a continuum. In addition to placing a fairly heavy constraint on the type of function that determines the outcome, this method does not allow for exceptions or special cases present in actual observations, which decision trees can easily accommodate.

One potential weakness of decision trees is that, although they are nonlinear classifiers, they are only capable of partitioning the data space along lines perpendicular to the attribute axes. However, this weakness is rarely a major issue, as multiple tests of the same attribute can be called in a single tree such that, given enough data, any partitioning of the space can be approximated with arbitrary precision. Further, since attributes can be defined as arbitrary nonlinear functions of simpler variables, nonlinearity in the data can be captured by the choice of axes as well as the partitioning of the space.

Another possible concern is decision trees' potential instability when data labels are perturbed (R. Li & Belford, 2002). Such instability is typically seen only in pathological cases, however. Furthermore, as much of the analysis described in this work involves the *collective* behavior of multiple models, any potential instability of individual models, however unlikely, is of lesser concern.

Decision Trees in Action. It will be helpful to build intuition by first looking at a simple example of what this technique does with data. Imagine that we have encountered a mysterious mechanical object that processes data according to some unknown rule. Data streams into the device, and out through a speaker in its side chirps a mechanical "yes" or "no," ostensibly in response, after each string of data that comes in. While we may be able to observe the device's output and monitor its input, we have no idea what rule the device is using to make its yes-or-no decision. That is something that we would like to know, so we make a record of what information is going in and what is coming out and hope that some analysis will reveal what the rule is.

Let us say that there are six streams of incoming data, which we label for convenience by the variables u, v, w, x, y, and z. We also track the "decision" variable by recording whether the device has responded to the data with a "yes" or a "no." We assume that the rule under the hood of this "robot" corresponds to a relationship between the input variables and the output, so we try to identify this rule by analyzing the data we have gathered using a decision-tree learning algorithm.

We made up a rule for the robot to follow and used the decision-tree algorithm `J48`, a Java implementation of Ross Quinlan's well-documented `C4.5` algorithm (Quinlan, 1993) running in the open-source platform Weka (Hall et al., 2009; Witten, Frank, & Hall, 2011), on 100 simulated observations to produce the tree shown in Figure 3 using the algorithm's default settings.[10] Several things are worth pointing out about this decision

[10]The `J48` algorithm has two options set by default in Weka. The first is a "confidence factor," set to 0.25, which has some bearing on the pruning of the tree. Some tuning of this parameter can lead to a better balance between training- and test-set classification performance, but as there is no *a priori* reason to set it to another value prior to examining the data, we left it at the default setting, which performs well in

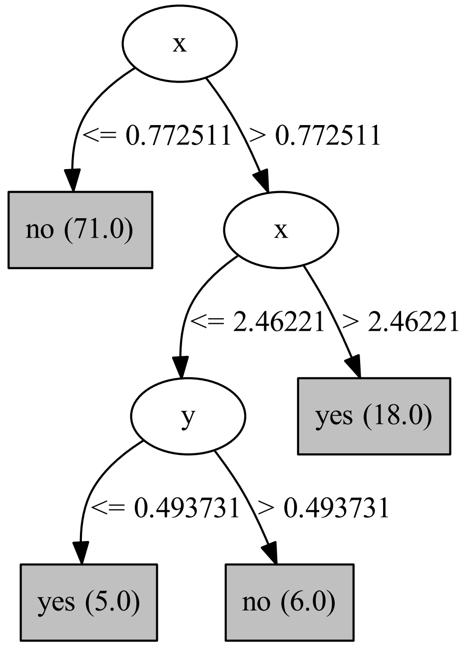


Figure 3. Decision tree generated from transformed raw "robot" data.

tree. First, it correctly classifies *all* of the data, and it does so *without* using the variables u, v, w, or z.[11] Second, it is able to immediately model the device's process on *new* data. That is, the very *description* of the pattern in the data implies a rule.

Using this tree's model to classify another, independent set of 100 observations of the robot results in a performance with an accuracy of 93 percent. It is clearly capturing a pattern in the data, and that pattern—which can be reasonably interpreted as a "signature" of the process generating the decisions—does not seem to involve the variables

practice. The second option is the minimum number of data instances that can be covered by a split in the tree, set to two by default. We manipulate this parameter as part of our analysis in a later section. The `J48` algorithm uses several pruning methods to prevent overfitting, including a technique based on the minimum description length (MDL) principle. Details can be found in Quinlan (1993) and Witten et al. (2011).

[11] This use of certain information to the exclusion of other variables deemed irrelevant is known as *feature selection*. A basic method of feature selection is part of the `C4.5/J48` algorithm, and it is another advantage of using the decision-tree technique.

u, v, w, or z.

While 93 percent accuracy is extraordinarily good, it is tempting to see whether an even more accurate model can be obtained by taking the two key variables identified by this tree, x and y, and creating a *new* data set by combining them in various simple functions. So, for instance, a new data set based on the original data can be created by getting rid of u, v, w, and z and computing the "new" variables x, y, x^2, y^2, xy, and x/y. Again using the `J48` algorithm on this transformed data set, one obtains the decision tree shown in Figure 4.

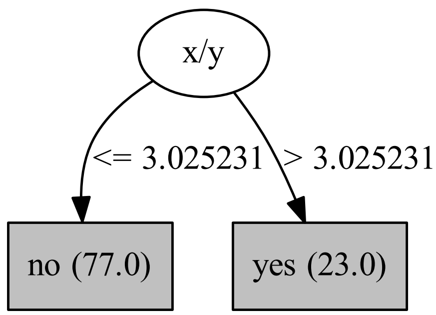


Figure 4. Decision tree generated from transformed artificial "robot" data.

It turns out that this decision tree, which chooses only one of the transformed data's attributes, x/y, and ignores everything else (including the single x and y attributes it used to create the first tree), reveals essentially the exact rule programmed into the device. That rule, artificially created for this example, is given by

$$D = \begin{cases} \text{yes} & x/y \geq \pi = 3.14159\ldots \\ \text{no} & \text{otherwise} \end{cases}. \tag{2}$$

To detect this rule in its correct form, it was necessary to include the variable combination x/y among the attributes in the second run of the algorithm.[12] This example illustrates,

[12]Note, however, that the rule given above is equivalent to saying "yes" whenever $x - \pi y \geq 0$ and "no"

though, that useful results are possible (as in the first stage of this example) even if the precise form of the variable of interest is not initially known. However, the algorithm must have the involved variables supplied *in some form*—and, at least initially, in a relatively simple form—in order to have a chance of detecting a useful signal in the data.[13] Although this example's data are noiseless for the purposes of illustration, meaning that the rule is always faithfully followed, the resulting model would be essentially the same even with noisy data.

A Decision Tree Analysis of Fajen and Devaney (2006)

In this section, we use decision tree analysis to analyze data from Experiments 1A and 1B of Fajen and Devaney (2006), the study described earlier in which subjects learned to perform an emergency braking task by timing the application of a fixed rate of deceleration so as to come to a stop as close as possible to a row of stop signs. Our aim was to use decision tree analysis to determine the optical variables upon which subjects relied and how the relied-upon variables changed with practice.

Methods

A brief summary of Fajen and Devaney. Before describing the analyses, we will summarize the relevant details of the methodology used in Experiments 1A and 1B of Fajen and Devaney (2006). In Experiment 1A, 20 subjects completed 10 blocks of 30 trials per block. Initial speed was a constant 10 m/s for all subjects, but target-sign radius varied

otherwise. In principle, this rule could have been detected on the first run through the data using the more general technique of *multivariate* or *oblique* decision trees (Alpaydin, 2010). On the other hand, this multivariate technique is most useful once feature selection has already taken place.

[13]This rule of thumb is reminiscent of what is known as the *sparsity-of-effects principle*, most often applied to regression models, which states that most real processes are dominated by the influence of *main effects* and low-order interactions of variables of interest, as opposed to more complicated higher-order functions of those variables (Montgomery, 2008).

from trial to trial, with its value (in meters) taken from the set

$$R \in \{0.150, 0.165, 0.195, 0.255, 0.375, 0.615\}.$$

Starting distances were chosen so that the initial retinal angle of the sign was $\theta_0 = 1.2$ degrees. In Experiment 1B, 16 subjects were run in 10 blocks of 30 trials per block. Sign radius was a constant 0.225 m for all subjects, but initial speed varied from trial to trial, with its value (in m/s) taken from the set

$$v_0 \in \{3.6585, 6.0, 8.8235, 11.5385, 13.6364, 15.0\}.$$

Again, starting distances were chosen so that the initial retinal angle of the sign was $\theta_0 = 1.2$ degrees.

To determine the role of global optic flow rate (GOFR), which specifies speed of self-motion, the presence of the textured ground plane was manipulated between subjects. That is, in both experiments, half of the subjects performed the task within an environment with a textured ground plane (the “Ground” condition), while the other half performed the task in an environment without this ground plane (the “Air” condition).

Decision Tree Analyses. We used the procedure illustrated in Figure 2 and a value of Δt equal to 90 ms to create two virtual data points for each trial—one representing the “do not brake” zone and the other representing the “brake” zone. This yielded a new data set composed of 600 data points per subject (10 blocks $\times$ 300 trials per block $\times$ two data points per trial). For each data point, we recorded the values of θ, $\dot{\theta}$, and velocity (v, which is proportional to GOFR) at the moment of brake initiation. From these variables, we calculated τ, $1/\tau$ and v/τ at brake initiation. Henceforth we will use v/τ rather than GOFR$\frac{\dot{\theta}}{\theta}$ (the label used by Fajen and Devaney) in the interest of brevity.

For the first set of analyses, we used the `J48` algorithm, again with its default settings, in Weka 3.7 (Hall et al., 2009) to create one decision tree from the transformed data per experimental block per subject, for a total of 360 trees for the 36 subjects in both parts of the experiment. Data sets from each block consisted of 60 observations (a pre-braking and post-braking observation from each of the 30 trials per block) of the optical variables, with each observation vector paired with one of two action classes, namely "do not brake" or "brake."

Results

Tree Size and Classification Rate. As illustrated in the box plots on the top row of Figure 5, trees were highly accurate in their description of the subject data, with an average correct classification rate of 89.1 percent (median 90).[14] The bottom row of Figure 5 shows the number of variable calls (or attribute tests), which provides a measure of the size or complexity of the trees; this is simply the number of nodes in the tree that create splits within the data. As the figure illustrates, the tree models were generally small, requiring an average of 2.21 attribute tests on the data (median 2).

Weka's implementation of the `J48` tree-learning algorithm produces "pessimistic" out-of-sample error estimates by default based on 10-fold cross-validation.[15] These estimates are not reported here for two main reasons: First, here we are interested in the models' compact *descriptions* of the training data as they may relate to control strategies. Second, we demonstrate out-of-sample performance more purely via the next-block prediction performance discussed below.

Over the course of an experiment, it is reasonable to expect that subjects will reduce their exploration of strategies as they learn the task. This suggests that deviation from a fixed strategy should decrease within each experimental block as the experiment goes on. Translated into modeling terms, each subject's behavior from within an experimental block should be better described by a *single* model as one progressively examines data from later in the experiment. That is, if bona fide strategies are being uncovered by the models, there should be a trend toward greater classifier accuracy versus block number, a measure of

[14] Reported accuracy is for within-sample performance. See discussion in the following paragraph.

[15] Here the "pessimism" takes the form of creating a confidence interval for the error rate and using its upper bound.

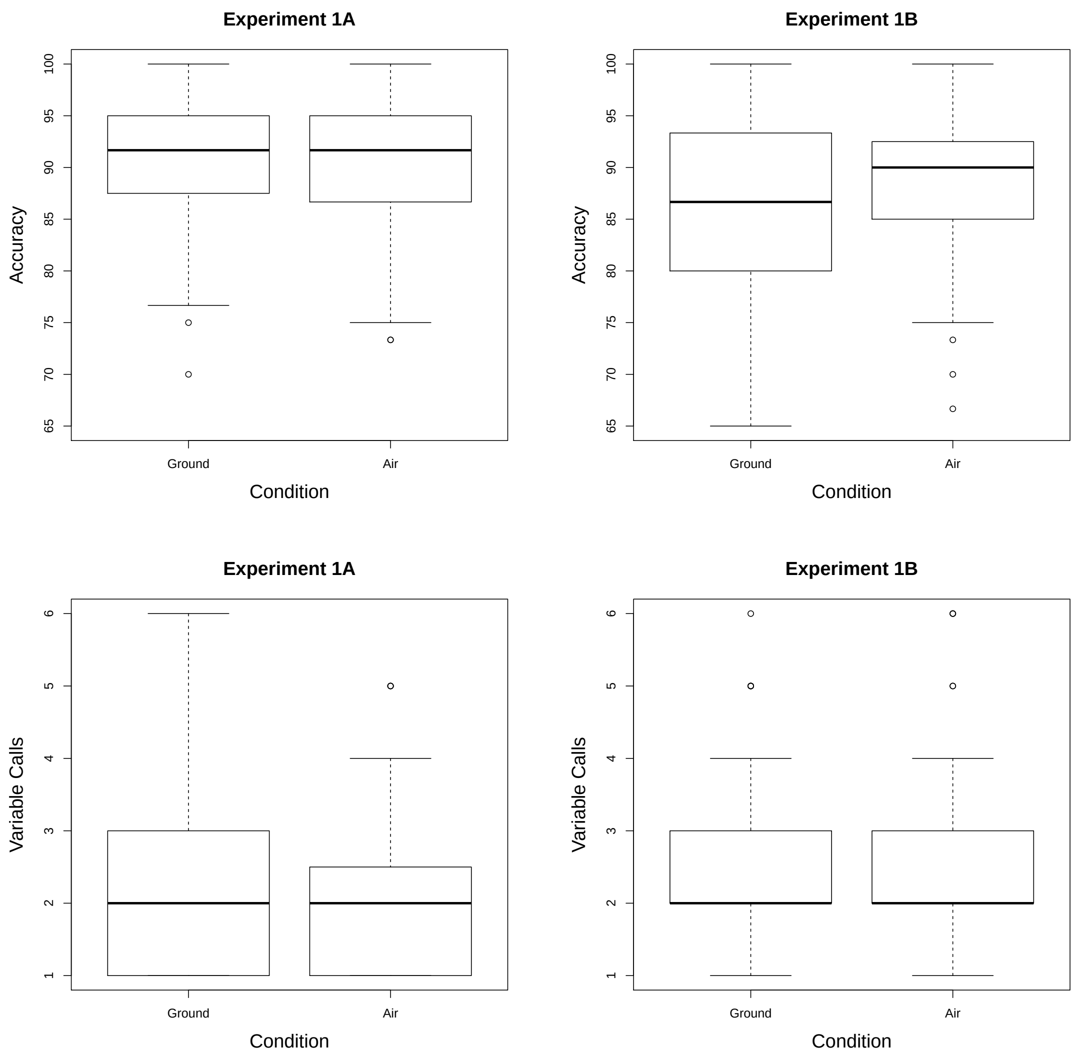


Figure 5. Box plots summarizing within-sample classifier accuracy (top) and number of attribute tests (bottom) by experiment and condition.

task-exposure time, since subjects will eventually settle on a strategy that can be described by a single model.

Indeed, a strong trend exists in the relationship between classifier accuracy and experimental block, which is illustrated in Figure 6A. There was a significant main effect of block on classifier accuracy on data from both Experiment 1A (Kruskal-Wallis $\chi^2(9) = 40.5912$, $p \ll 0.001$) and Experiment 1B ($\chi^2(9) = 21.0302$, $p < 0.05$). This manifested as a significant correlation between block and classifier accuracy in both experiments (Kendall's $\tau \approx 0.313$ and 0.251, respectively, both $p \ll 0.001$).

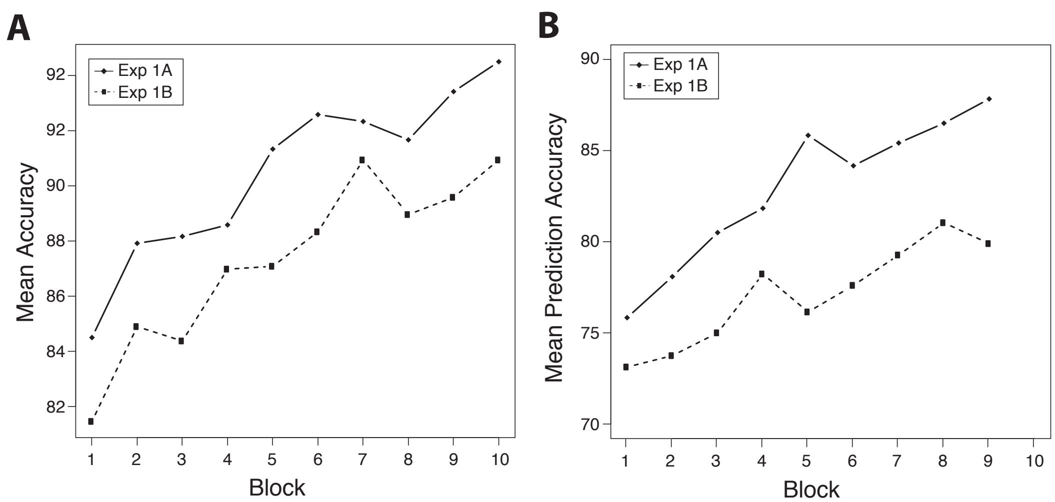


Figure 6. Mean classifier accuracy (A) versus experimental block and block-to-block prediction accuracy (B) versus experimental block.

It is also reasonable to expect that as exploration is reduced, the strategy in place in one block will be at least *similar to* the strategy used in the following block, even if subject strategies change somewhat over the course of the experiment. If one then uses a model built on one block's data to classify the data from the *following* block, one should observe a trend toward greater accuracy versus experimental block. This trend is indeed present, as illustrated in Figure 6B. Note that this also speaks to the models' performance on out-of-sample data, with the caveat that data coming from a later block may be generated by a strategy shaped by greater exposure to the task.

Finally, one should expect a model capturing genuine patterns in the data to perform on next-block data at a level of accuracy above chance. For the Fajen and Devaney (2006) data, chance performance of a classifier would be centered on a value of $p_0 = 0.5$, since the data are evenly split between classes. Using a standard normal approximation (Crow, Davis, & Maxfield, 1960), with

$$z = \frac{r - np_0 - 1/2}{\sqrt{np_0(1-p_0)}}, \tag{3}$$

where r is the number of data points correctly classified, a one-tailed test at the 0.95 significance level suggests that models relying on chance performance would achieve a better than 0.6145 classification rate only about five percent of the time. This level of performance on next-block prediction accuracy was exceeded in all but one case out of 180 (99.4 percent) in the Experiment 1A data set and all but six cases out of 144 (95.8 percent) in the Experiment 1B data.

Although the trends illustrated in Figure 6 are similar, they speak to slightly different phenomena. The first trend indicates that patterns detected in subject data are increasingly *consistent within blocks* as the experiment progresses. The second trend indicates that the patterns are increasingly *persistent between blocks* as the experiment progresses.

Accuracy in describing a data set is highly desirable in any model, but it is not the only criterion on which tree models should be rated. Some consideration must be given to the possibility that this accuracy comes at too high a price if arbitrarily complex models are built on data sets in which no genuinely useful patterns exist. What follows should provide some assurances against this possibility.

Decision trees reorganize data sets by sorting them according to values of their attributes. This partitions the class values into several bins, with each bin corresponding to a leaf of the tree (see Figure 7). If no genuine patterns linking input to output exist in the data, then any tree built using those data would either not significantly change the distribution of class values along the tree's branches or would do so only by growing to

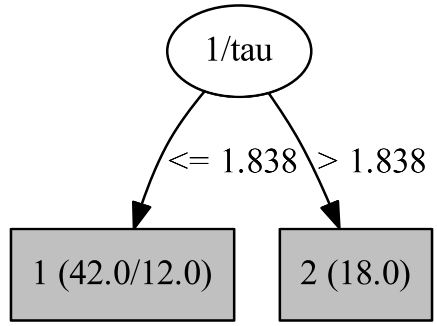


Figure 7. Model output for Subject 10 in Block 10 of Experiment 1A. The decision tree partitions the class data into two bins, the first containing 42 elements (with 12 misclassifications) and the second containing 18 elements (with no misclassifications) after sorting on $1/\tau$.

have almost as many leaves as data points.

One can test both possibilities by computing a χ^2 goodness-of-fit statistic for each leaf and summing them together, getting

$$\sum_{i=1}^{L} \frac{(N_i - 2m_i)^2}{N_i} \sim \chi^2(L), \tag{4}$$

where L is the number of leaves in the tree, N_i is the number of instances covered by leaf i, and m_i is the number of instances misclassified in leaf i.[16] (For the tree in Figure 7, $L = 2$, $N_1 = 42$, $m_1 = 12$, $N_2 = 18$, and $m_2 = 0$.) This measure not only rewards a model for its accuracy but also penalizes it for making too many tests on the attribute values, which is directly related to the number of leaves in the tree. For the tree in Figure 7, for instance,

[16]This expression takes this form because for each i,

$$\frac{(N_i/2 - m_i)^2}{N_i/2} + \frac{(N_i/2 - (N_i - m_i))^2}{N_i/2} = \frac{(N_i - 2m_i)^2}{N_i} \sim \chi^2(1).$$

A conservative statistic, which applies the *Yates correction*, can be used when $N/(2L) < 10$, and it is given by $\sum_{i=1}^{L}(|N_i - 2m_i| - 1)^2/N_i$.

Table 1

Significance table for "suspect" models. Yates-corrected (YC) values are given for larger trees. The left three columns describe the source data for training the models.

Exp	Subject	Block	Accuracy	Leaves (L)	χ^2 (YC)	$\chi^2_{0.95}(L)$
1A	7	1	70	2	15	5.99
1A	7	3	93.3	7	48.37 (38.67)	14.07
1B	5	3	65	2	10.59	5.99
1B	5	10	95	7	51.11 (38.02)	14.07
1B	6	9	96.7	7	53.5 (42.58)	14.07
1B	6	10	96.7	7	38.31 (29.11)	14.07
1B	14	5	98.3	7	56.5 (44.89)	14.07

one calculates a $\chi^2(2)$ value of 25.71, which is well above even the most conservative critical value. Strictly speaking, this statistic tells us that the new model gives us a significantly *different* picture of the class distribution than the null model does. But, since we also know the accuracy is greater, we can reasonably conclude that the model is significantly *better*.

Rather than present results for separate calculations of this statistic for all 360 models, we will instead choose models from each experiment whose characteristics make them the most "suspect." Namely, these will be models with the lowest classification accuracy as well as those with the most leaves, even if their classification accuracy is high. The statistics for these models, all of which exceed their respective critical values, are presented in Table 1. This analysis reassures us that these models do not achieve their high accuracy by producing arbitrarily large, uninterpretable trees.

Simulating Subject Behavior. A major test of the effectiveness of the tree models is the accuracy with which they simulate subject behavior. This was evaluated by creating a simulated environment (in MATLAB) that mimicked the conditions of the original experiments. The simulation took the form of having the subject tree models classify each point in the simulation data set as one of two classes: “do not brake” and “brake.” The first point to be classified as “brake” in each set was taken to be the point at which the model applied the brake. The models’ braking decisions were then recorded (in spatial coordinates) and averaged across models in a manner analogous to the across-subjects approach used by Fajen and Devaney (2006). The averaged simulation results are shown in Figure 8 (solid lines) along with the data from Fajen and Devaney (dashed lines). The figure clearly shows that the averaged behavior of the models closely matches the characteristics of the averaged subject behavior from the original study.

Interpreting Trees as Structural Models. Having established that the decision-tree models are capable of capturing the key aspects of subjects’ behavior, we consider how to analyze these models to glean insight into the optical variables upon which subjects relied and how the relied-upon variables changed throughout the experiment.

Of particular interest are the specific *initial variable calls* that the models make. The initial variable call is the top-node variable in a tree model, the variable first queried by the model to split the data into subsets, with the specific query chosen to maximally reduce the entropy of the resulting data subsets relative to the original data set. (Informally, it can be thought of as the model’s best opening gambit in a game of “20 questions.”) Recall that the brake training data were evenly split between “brake” and “do not brake” examples. The initial variable call corresponds to the question we can ask about a value of one of the data’s attributes that, once answered, will give us the best overall chance of guessing the correct outcome (“brake”/“do not brake”) in one shot. As such, it represents the data’s “dominant” underlying pattern in the guise of its most informative variable or attribute.

If the strategies behind the subject data shift during the course of the experiment—such as the transition from the $\dot{\theta}$ strategy to the $1/\tau$ strategy suggested by Fajen and Devaney (2006) in Experiment 1A—then this shift should also be evident in the record of model variable calls. Specifically, this shift will take the form of a majority of tree models built on early-block data (with each model representing an individual subject)

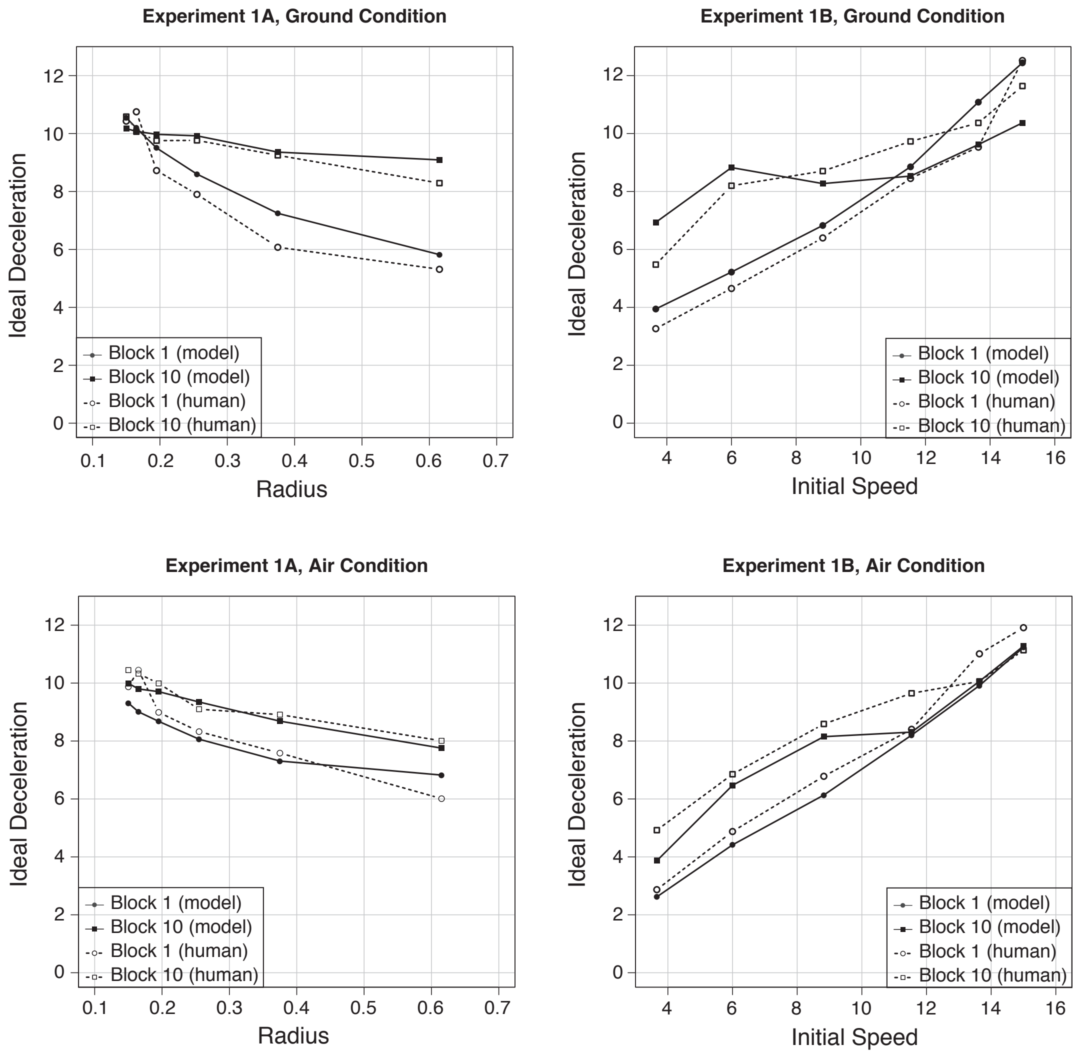


Figure 8. Simulation results showing mean ideal deceleration at brake onset as a function of sign radius in Experiment 1A and as a function of initial speed in Experiment 1B. Original data from the Fajen and Devaney (2006) study are reproduced as dashed lines.

calling $\dot{\theta}$ initially (that is, having $\dot{\theta}$ as the top node) and a majority of late-block models initially calling $1/\tau$ instead.

The distribution of initial calls to specific variables by experiment and block is illustrated in Figure 9. As one can see, in Experiment 1A, $1/\tau$ emerges as a clear winner over the course of the experiment in both conditions, serving as the top-node variable in almost every individual-subject model built on late-block data, while serving in that role in only 20 to 30 percent of the models from early-block data, whose models were dominated by $\dot{\theta}$. Further, note that this transition appears along a continuum over the course of the experiment. This is exactly what one would expect if individual subjects shifted their strategies at different times during the experiment.

The picture is less obvious in the Ground condition of Experiment 1B, where there appears to be more competition among variables for the models' initial tests. This does not necessarily imply, however, that the models failed to identify strategies that accurately represented the subject behavior. Rather, it could be that subjects' braking decisions were often well described in terms of *multiple* strategies defined over several variables, although the $1/\tau$ strategy often predominates. Again, this is largely consistent with the Fajen and Devaney analysis. The situation is a bit more straightforward in the Air condition of Experiment 1B, where the $1/\tau$ strategy again predominates.

The reader may note from Figure 9 that in the Air condition of Experiment 1B the attribute v/τ was occasionally called even though velocity-scaled information in the form of GOFR was ostensibly unavailable. Although it may ultimately be reasonable to refrain from supplying information to a classifier that subjects could not have used during the actual experiment, there is at least some benefit in initially providing it. First, supplying supposedly unused information, at least initially, can provide a "reality check" for the models that are produced by the algorithm. Second, in certain experimental settings, it may be possible that subjects discovered a source of information that the experimenter had not identified (e.g., a source of information about speed of self-motion that was available

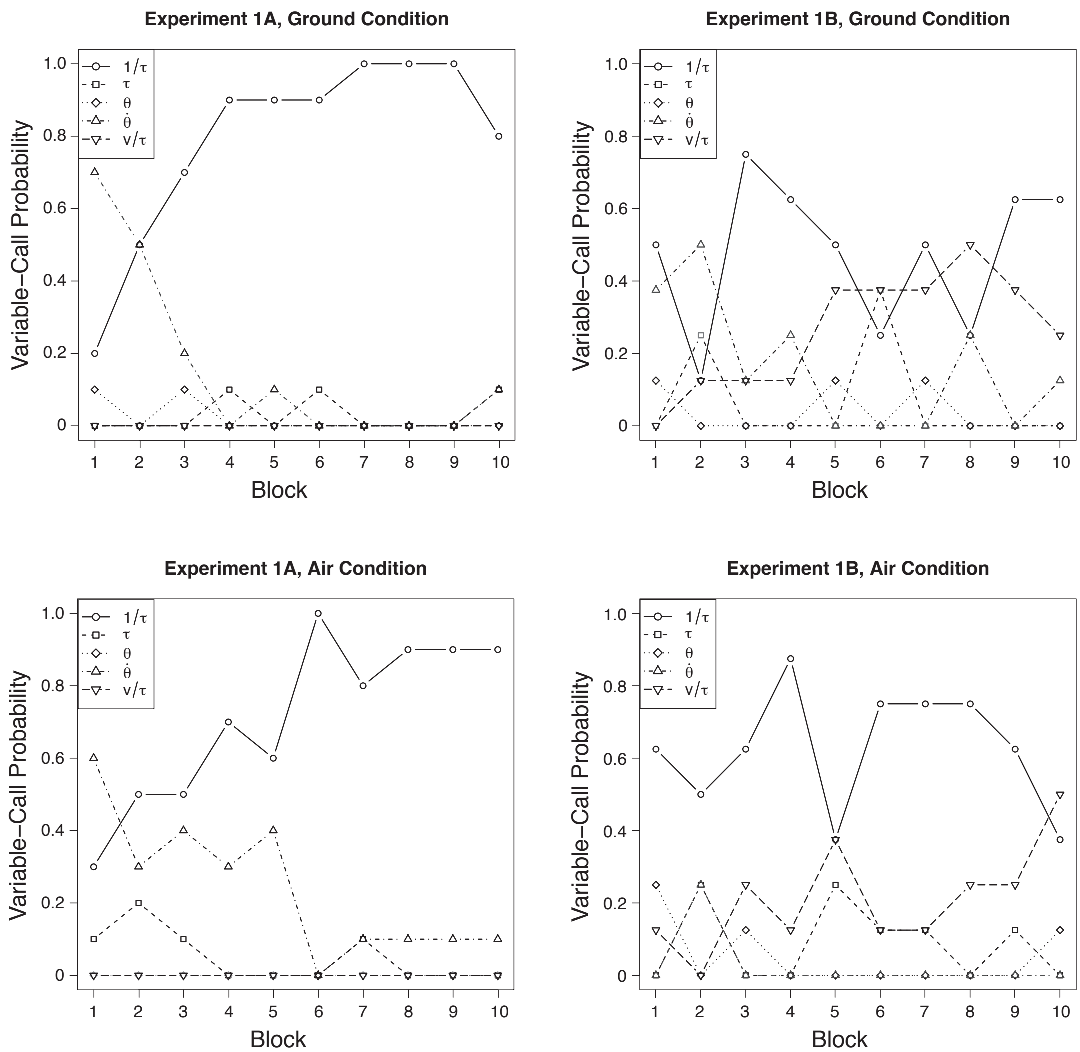


Figure 9. Distribution of top-node variables (initial variable calls) of individual-subject models grouped by experiment and condition.

even in the Air condition).

In all cases tested drawn from the Air condition, data sets without velocity-scaled attributes were still capable of producing models of subject data of comparable accuracy to those formed from data sets with velocity-scaled attributes. As a result, the ensemble models for data gathered from the Air condition, discussed below, were built from models in which the velocity-scaled information was removed.

Producing and Interpreting Summary Models. It would be helpful to form a single representation of the results of multiple models, much as one would average quantitative data to form a summary statistic. Such summaries can then be grouped by various factors and examined for potentially meaningful differences. Since the models produced so far are graphical structures, this presents an interesting challenge. One has a choice of how to go about making such summary models. The following are several options:

1. The data used to build the original models can be combined into a single (very large) set, and a single model can be built on this set.

2. The individual subject data can be averaged by condition and action class and then combined into a single (moderately large) training set for modeling.

3. The individual subject data can be combined first and *then* averaged by initial condition and action class, forming a single (relatively small) training set for modeling.

4. The collective behavior of the models can be captured through an *ensemble* model.

In every case but the last, one would be discarding the models that had already been built and analyzed. In the ensemble option, though, every model built from subject data is used to classify a large test set, made up of unlabeled, simulated data covering *all* of the experimental conditions (see Figure 10). Each model's classification of a data point is essentially a "vote" for a particular category or class of action (e.g. "brake" or "do not brake"). When the ensemble's votes are considered together, the category getting the most votes for each data point is assigned as its label. This produces a *new* training set upon which a single model can be built.

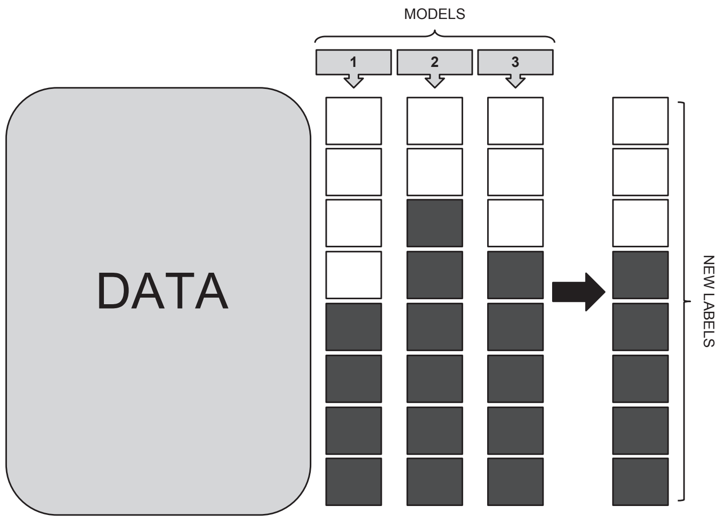
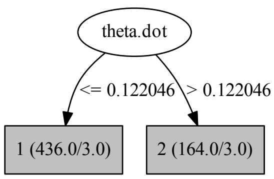


Figure 10. A schematic representation of ensemble averaging.

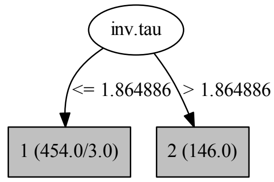
We will focus on this method, since we are interested in summarizing the *models*, which have already demonstrated their ability to simulate the characteristics of subject behavior. Ensemble models for the beginning, middle, and end blocks of Experiment 1A are shown in Figure 11. The ensemble summaries for Experiment 1B are shown in Figure 12. Note that the general strategic trends in these models reflect those discussed in the original Fajen and Devaney (2006) paper. Namely, in Experiment 1A, initial strategies appeared to be consistent with a $\dot{\theta}$ strategy before moving toward a $1/\tau$ strategy as the experiment progressed; in Experiment 1B, subject behavior initially appeared consistent with a $1/\tau$ strategy before moving toward a policy more closely resembling a v/τ strategy.[17]

Using decision trees to understand human behavior. In this section, we discuss how decision tree modeling and analysis allow us to go beyond the standard approach and gain deeper insight into subjects' control strategies. Let us reconsider the performance of human subjects in the Ground condition in Experiment 1B (see Figure 1B). Recall that performance in Block 1 was consistent with a $1/\tau$ strategy, with a significant

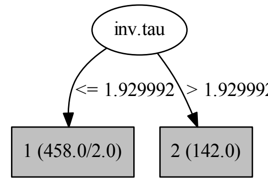
[17]In the case of the final block of Experiment 1B's Air condition, in which v/τ was not available, the result was a more complicated, "hybrid" strategy.



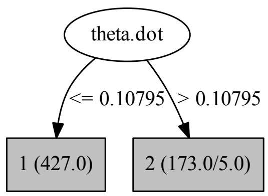
(a) Block 1 (Ground)



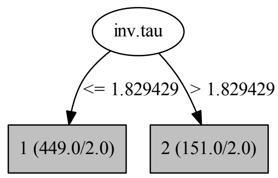
(b) Block 5 (Ground)



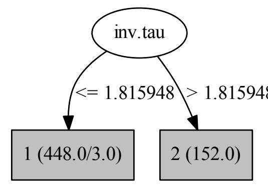
(c) Block 10 (Ground)



(d) Block 1 (Air)



(e) Block 5 (Air)



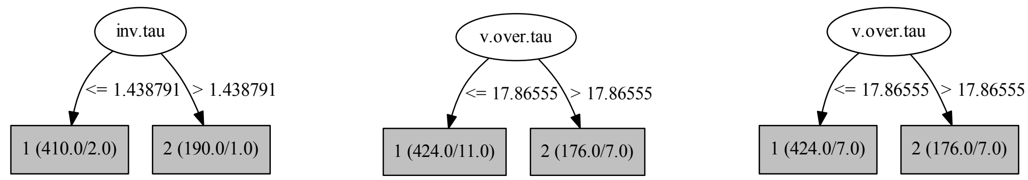
(f) Block 10 (Air)

Figure 11. Experiment 1A ensemble models.

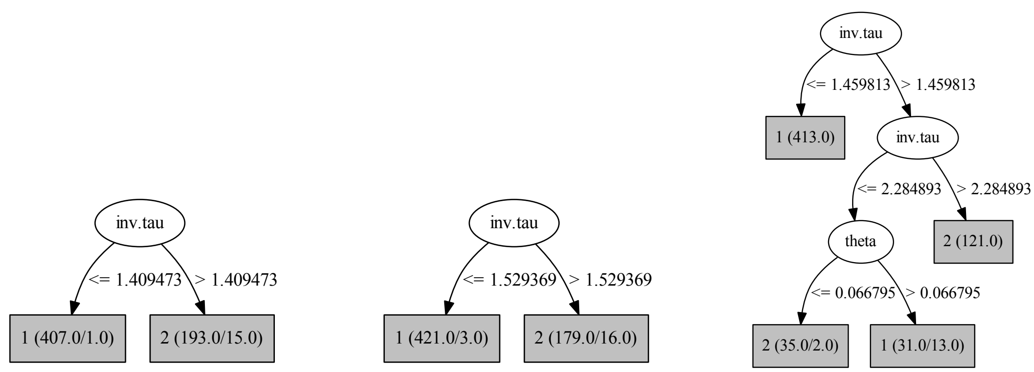
early braking bias for all but the fastest initial speeds. This bias was reduced with practice, suggesting that subjects learned to use a source of information that was more effective than $1/\tau$. However, the bias persisted even through the tenth block, indicating that the variable upon which subjects relied was not the optical invariant (v/τ).

If subjects did not use $1/\tau$ or v/τ, then what source of information did they use in Block 10? This question is difficult to answer if one's candidate sources of information are limited to optical variables that can be expressed using algebraic combinations of optical primitives. However, because decision tree models can represent more complex combinations of optical primitives, they allow us to better understand what subjects are doing when their behavior falls in between the predictions of optical variables identified using the standard approach.

To illustrate the process, we created three versions of the ensemble model for Block 10 in the Ground condition of Experiment 1B (see Figure 13). The different versions were



(a) Block 1 (Ground) (b) Block 5 (Ground) (c) Block 10 (Ground)



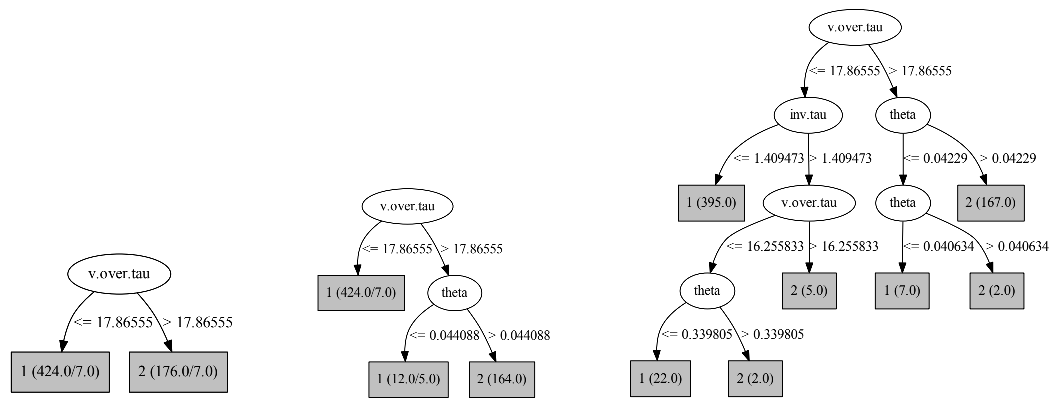
(d) Block 1 (Air) (e) Block 5 (Air) (f) Block 10 (Air)

Figure 12. Experiment 1B ensemble models.

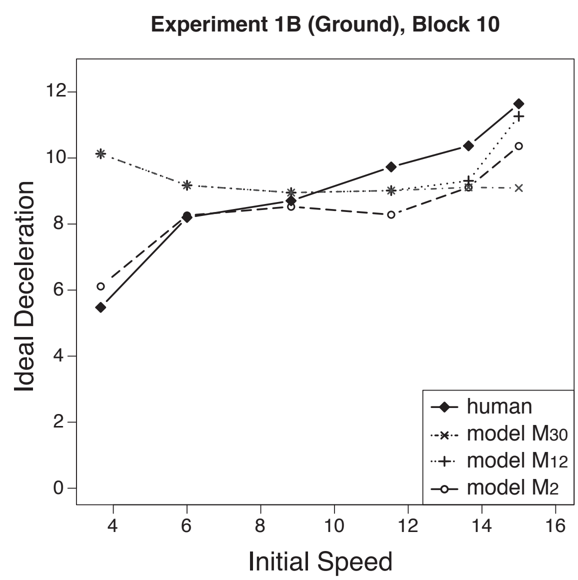
created by varying the minimum number of instances (M) that must be covered by a leaf in order for that leaf to be included in the model. Thus, models created with smaller values of M tend to be more complex and have more nodes. We tested the following three values of M: 30 (Figure 13a), 12 (Figure 13b), and 2 (Figure 13c).

The model that was created using $M = 30$ (which we refer to as the M_{30} model) has a single node (v/τ). When we simulate this model using the same conditions that were used in Experiment 1B, braking is initiated at roughly the same value of ideal deceleration across the range of initial speeds (see $\times$ markers in Figure 13d). This is not surprising because v/τ invariantly specifies ideal deceleration across variations in initial speed.

Although the M_{30} model performs the task quite well, its behavior deviates from human behavior (see closed diamond markers in Figure 13d), especially in the slowest and



(a) Model M_{30} (b) Model M_{12} (c) Model M_2



(d) Model Simulations and Subject Behavior

Figure 13. Summary models for the Ground condition of Experiment 1B. Here, M denotes the minimum instances per leaf.

fastest initial speed conditions. With the addition of a second node that follows the test of v/τ with a test of θ, the M_{12} model better captures the bias in the human data to brake too late in the fastest initial speed condition (see + markers in Figure 13d). Further decreasing M down to 2 allows the model to capture both the early braking bias at slow initial speeds and the late braking bias at fast initial speeds (see $\circ$ markers in Figure 13d).

The M_2 model is informative not because it fits the human data more closely. (After all, the better fit is achieved only by a significant increase in model complexity.) Rather, by comparing the structure of the M_2 model to that of the M_{30} model, we can begin to answer the question posed earlier about what source of information subjects relied upon in Block 10.[18] Both models initially call v/τ, which invariantly specifies ideal deceleration. However, the M_2 model includes five additional nodes, three of which call θ. This suggests that the bias in the human data may be due to a residual influence of the optical angle of the stop signs, but only for values of initial speed at the extremes of the range. In other words, subjects learned to rely on the optical invariant (v/τ) for initial speeds in the middle of the range but were influenced by θ as well when initial speed was very slow or very fast.

This analysis illustrates one of the ways in which decision tree models can be used to go beyond previous attempts to identify the optical variables upon which subjects rely. Because the approach adopted by Fajen and Devaney (2006) was limited to consideration of a small set of possible optical variables, the best they could do was to conclude that behavior in Block 10 was similar but not identical to a v/τ strategy. With the ability to consider a much wider range of possible informational variables, decision tree analysis allows us to understand how subjects' control strategies deviated from the v/τ strategy and under what conditions.

[18] Furthermore, we can use this technique to formulate predictions to test in follow-up experiments. See the discussion.

Discussion

In this study, we introduced a new technique for uncovering visual control strategies based on human data. The technique uses decision tree learning algorithms to create tree-like structures that model the control strategy used by the subject when the data were generated.

The trees generated by this process were generally compact yet captured human behavior quite well. We also introduced a technique for creating ensemble models that effectively "average" together multiple models generated from individual subject data, creating a compact summary model. Lastly, we showed how decision tree models allow researchers to go beyond the standard approach and gain deeper insight into subjects' control strategies.

Benefits of decision trees

Decision tree learning has a number of benefits that we believe will appeal to a wide range of researchers, not just those interested in visual control and perceptual attunement. First, decision trees allow researchers to consider more complex sources of perceptual information. This is especially true if one envisions the modeling effort as a multistage process, with the initial stage serving to eliminate variables that have little apparent predictive utility and later stages introducing novel combinations and functions of the remaining variables and allowing the tree-learning algorithm to produce models from this expanded feature set. Second, decision tree models can be easily integrated into computational simulations under conditions that are similar to those used in the experiment from which the data were generated. Third, decision trees have the advantage of being easy to interpret, allowing researchers to form considerable intuition about the modeled process at a glance.

Decision trees have other properties that were not specifically demonstrated in this study but could be explored in future work.

Capturing learning as a gradual attunement process. Decision tree models may provide a useful framework for understanding the changes that take place during perceptual attunement. When observers transition from one optical variable (e.g., $1/\tau$) to another optical variable (e.g., v/τ) with practice, the transition does not necessarily take place instantaneously. That is, an observer does not necessarily rely on $1/\tau$ at one moment and v/τ at the next moment; the attunement process may be more gradual.

For example, the process of transitioning from attunement to $1/\tau$ to attunement to v/τ could begin with v influencing behavior for a narrow range of conditions—most likely, those that were encountered during practice with feedback. A decision tree model could capture this with the addition of a new node that gives rise to a v/τ-like strategy under some conditions and a $1/\tau$-like strategy in other conditions. As the observer practices the task and receives performance feedback across a wider range of conditions, and the v/τ strategy generalizes, the structure of the decision tree model will continue to change to capture this generalization. By analyzing changes in the structure of the model with practice, researchers can gain new insights into how attunement changes over time. Eventually, if the observer becomes attuned to v/τ across the entire range of conditions, the model would have a single v/τ node.

Thus, when using decision tree analysis, one is not forced to think about perceptual attunement as a process of sudden change (although decision tree models are certainly capable of capturing sudden changes in attunement when they occur). Decision tree analysis provides a natural framework within which to understand perceptual attunement as a process of gradual change.

The question of attunement has been considered in some depth by Jacobs, Michaels, and colleagues. Of particular interest in the context of the ecological theory of perception is their idea that perceptual attunement (or *education of attention*) can be viewed as a path through an *information space* and that key "information for learning" is available to the organism that essentially prescribes that path, allowing for a phenomenon they refer to as *direct learning* (Jacobs & Michaels, 2007).[19]

Of course, before the *researcher* can detect that path, he or she must first identify the variables of interest that form the axes of the traversed space. This presents a problem when the choice of variables is not obvious, and it highlights a strength of the method outlined in this paper, which can quickly wade through a sea of candidates and identify the variables most likely to have real bearing on the behavior under study. In this way, our method can be viewed as a useful complementary tool for researchers pursuing the direct-learning/information-space approach.

Generating new predictions. Decision tree models can also be used to generate both quantitative and qualitative predictions about behavior in novel situations. These predictions could then be tested in a follow-up experiment.

For example, in the transition from Model M_2 to Model M_{12} shown in Figure 13, several tests on the θ attribute are eliminated and collapsed into their parent nodes. The result, as shown in simulations, is the elimination of an early-braking bias at the low end of the initial-speed scale, while the rest of the behavior is essentially unchanged. As the models become simpler, eliminating the test on θ removes a late-braking bias at the high end of the initial-speed scale when the model is run as a simulation of braking behavior.

Although sign radius and initial speed were not varied within the same experiment in Experiment 1, this apparent sensitivity to θ can guide expectations for what might be observed in an experiment in which both factors are varied simultaneously. Specifically, actual subject behavior—as described by the most complex summary model—differs from a pure, unbiased v/τ strategy only in an apparent sensitivity to the optical angle of the target at the low and high ends of the speed scale.

In Figure 13d the simulated braking behavior remains stable in the middle ranges of speed against removal of the tests on the θ attribute. Examining the rules prescribed by

[19] A different interpretation of the concept of an information space is employed in the context of continuous control in Weber (2013).

the largest tree (M_2), one can translate this apparent size bias into the following prescription: "Brake for signs that loom larger, but only at the fastest and slowest speeds." In particular, the tendency appears to be for larger (and therefore larger-looming) signs to cause *late* braking at the highest speeds and *early* braking at the lowest speeds.

Fajen and Devaney performed a follow-up experiment in which sign size and initial speed were varied within the same block. One motivation for this second experiment was their belief that the relatively unremarkable differences between environmental (i.e. ground vs. air) conditions in Experiment 1B may have been due to the limited range of conditions created by varying only one factor at a time (Fajen & Devaney, 2006). Using the methods outlined in this work, a very *specific* effect of sign size can be hypothesized for this design using only data gathered from an experiment in which sign size did not vary at all.

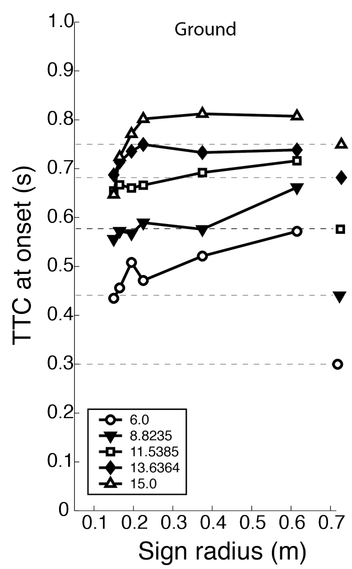


Figure 14. Time to contact (TTC) at brake initiation versus sign radius, grouped by initial speed. Figure 11(a) from Fajen and Devaney (2006).

A thorough analysis of this follow-up experiment will not be made in this paper, but data from this second experiment are consistent with the sign-size prediction made above. Figure 14 shows the time to contact (TTC) at brake initiation versus sign radius for the ground condition of their Experiment 2. The data have been grouped by initial speed, and one can clearly see a tendency for late braking (lower TTC values) at the highest initial speeds when sign radii are small and a tendency for early braking (higher TTC values) at the lowest speeds when sign radii are large. The TTC values are fairly stable, though, over all sign radii for the intermediate speed (11.5385 m/s).

This observation is entirely consistent with the prediction arising from the analysis of the models in Figure 13. It is not possible to explain this observation with the three hypothesized strategies tested in the original study. This is but one example of how the data-mining approach can inform behavioral predictions that can be explicitly tested in new experiments. This example illustrates how this analytical technique is best considered a complement to—and not a replacement of—careful experimental methodology.

Caveats and possibilities

While the decision tree method allows for considerable freedom relative to the standard hypothesis-driven approach, it is not free from assumptions. In order for a model to have a chance at capturing a control law, it is essential that the optical variables being employed are somewhere among the variables supplied to the learning algorithm. As we remarked earlier, it is not absolutely necessary that these variables be tracked from the start in their final functional form, but it is key that they be part of the set of variables under consideration.

If one includes too many variables in an exploratory analysis, one runs the risk of creating data that is *underspecified*, meaning that the number of observed outcomes is not sufficiently large relative to the number of variables being considered to explain them. Still, judicious selection of variables being considered at the beginning of one's study combined with the analytical and ensemble-averaging methods discussed in this paper should provide protection against this problem.

The task used to illustrate the methods in this paper resulted in a *discrete* behavior, namely the binary choice of whether to execute a single braking action. This does not mean, however, that this method cannot also be applied to behavior more commonly considered in the continuous domain.[20] Here some judgment on the part of the researcher is necessary as to which aspect of the behavior is the most interesting focus of analysis. For certain continuous behaviors, the *initiation* of the action may be of the most interest, while for others it may be the timing of a maximal or threshold event. Continuous actions can also be discretized and decomposed into stereotyped stages. This expands the problem from the binary classification problem we considered here to a multiclass classification problem, but the method remains essentially the same.

Although our focus in the present study was on visual information, there is no reason why decision tree modeling could not be used to identify information in other sensory arrays. In fact, this approach may turn out to be even more fruitful in the study of haptics and audition, since considerably less is known about how to describe informational variables in these other modalities. Likewise, decision tree modeling may facilitate attempts to identify information that is defined across multiple sensory arrays.

It is our hope that the method we describe here will appeal to a wide range of researchers and that the broader mining of experimental data will afford us additional glimpses under the hood of human behavior.

[20] A different treatment of mining control laws from continuous data can be found in Weber (2013).

References

Alpaydin, E. (2010). *Introduction to machine learning* (Second ed.). Cambridge, MA: MIT Press.

American Psychological Association. (2014). Exploratory data mining in behavioral research.. Retrieved from `http://www.apa.org/science/resources/ati/data-mining.aspx` (Last accessed July 28, 2014.)

Baayen, R. H., & Cutler, A. (2005). Data mining at the intersection of psychology and linguistics. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 69–83). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chapman, S. (1968). Catching a baseball. *American Journal of Physics*, *36*, 368–370.

Chattamvelli, R. (2011). *Data mining algorithms.* Oxford: Alpha Science.

Crow, E., Davis, F., & Maxfield, M. (1960). *Statistics manual.* New York: Dover.

Downing, D., Covington, M., & Covington, M. (Eds.). (2000). *Dictionary of computer and internet terms* (Seventh ed.). Hauppauge, NY: Barron's Educational Series.

Fajen, B. R. (2005a). Calibration, information, and control strategies for braking to avoid a collision. *Journal of Experimental Psychology: Human Perception and Performance*, *31(3)*, 480-501.

Fajen, B. R. (2005b). Perceiving possibilities for action: On the necessity of calibration and perceptual learning for the visual guidance of action. *Perception*, *34*(6), 741–755.

Fajen, B. R. (2005c). The scaling of information to action in visually guided braking. *Journal of Experimental Psychology: Human Perception and Performance*, *31(5)*, 1107-1123.

Fajen, B. R. (2007). Rapid recalibration based on optic flow in visually guided action. *Experimental Br*, *183*, 61-74.

Fajen, B. R. (2008a). Learning novel mappings from optic flow to the control of action. *Journal of Vision*, *8*(11), 1–12.

Fajen, B. R. (2008b). Perceptual learning and the visual control of braking. *Perception & Psychophysics*, *70*(6), 1117–1129.

Fajen, B. R. (2013). Guiding locomotion in complex dynamic environments. *Frontiers in Behavioral Neuroscience*, *7:85*, 1-15.

Fajen, B. R., & Devaney, M. (2006). Learning to control collisions: The role of perceptual attunement and action boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 300–313.

Fajen, B. R., & Warren, W. (2004). Visual guidance of intercepting a moving target on foot. *Perception*, *33*(6), 689–715.

Gibson, J. (1986). *The ecological approach to visual perception.* London: Psychology Press.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*(1), 10–18.

Jacobs, D. M., & Michaels, C. F. (2007). Direct learning. *Ecological Psychology*, *19*(4), 321–349.

Lee, D. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, *5*, 437–459.

Lenoir, M., Musch, E., Thiery, E., & Savesbergh, G. J. P. (2002). Rate of change of angular bearing as the relevant property in a horizontal interception task during locomotion. *Journal of Motor Behavior*, *34*, 385-401.

Li, R., & Belford, G. (2002). Instability of decision tree classification algorithms. In *Proceedings of the eighth* ACM SIGKDD *international conference on knowledge discovery and data mining* (pp. 570–575).

Li, W., Saunders, J., & Li, L. (2009). Recruitment of a novel cue for active control depends on control dynamics. *Journal of Vision*, *9*(10), 1–11.

Montgomery, D. C. (2008). *Design and analysis of experiments* (Seventh ed.). New York: John Wiley & Sons.

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about* *data mining and data-analytic thinking.* Sebastopol, CA: O'Reilly.

Psychonomic Society. (2012). Psychonomics without experiments? Discovering psychological principles by mining large data sets. In *Abstracts of the psychonomic society* (Vol. 17). Retrieved from `http://www.psychonomic.org/pdfs/PS_2012_Abstract_Book_WEB.pdf` (Last accessed March 15, 2013.)

Quinlan, J. (1993). *C4.5: Programs for machine learning.* San Francisco: Morgan Kaufmann.

Smith, M., Flach, J., Dittman, S., & Stanard, T. (2001). Monocular optical constraints on collision control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(2), 395–410.

Stanard, T., Flach, J., Smith, M., & Warren, R. (2012). Learning to avoid collisions: A functional state space approach. *Ecological Ps*, *24*, 4.

Theodoridis, S., & Koutroumbas, K. (2009). *Pattern recognition* (Fourth ed.). Burlington, MA: Academic Press.

Warren, W. (1988). Action modes and laws of control for the visual guidance of action. In O. Meijer & K. Roth (Eds.), *Complex movement behaviour: The motor-action controversy* (pp. 339–380). Elsevier.

Warren, W. (1998). Visually controlled locomotion: 40 years later. *Ecological Psychology*, *10*, 177–219.

Weber, R. (2013). *Discovering optical control strategies: A data-mining approach.* Unpublished doctoral dissertation, Rensselaer Polytechnic Institute.

Witten, I., Frank, E., & Hall, M. (2011). *Data mining: Practical machine learning tools and techniques* (Third ed.). New York, NY: Morgan Kaufmann.

Yilmaz, E., & Warren, W. (1995). Visual control of braking: A test of the $\dot{\tau}$ hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 996–1014.