This is not a final draft. Final draft available from:

An Empirical Illustration and Formalization of the Theory of Direct Learning:

The Muscle-Based Perception of Kinetic Properties

David M. Jacobs\textsuperscript{1,2}, Paula L. Silva\textsuperscript{2}, and Juan Calvo\textsuperscript{3}

\textsuperscript{1} Facultad de Psicología, Universidad Autónoma de Madrid
\textsuperscript{2} Center for the Ecological Study of Perception and Action, University of Connecticut
\textsuperscript{3} Departamento de Matemática Aplicada, Universidad de Granada

Please send correspondence to:

David M. Jacobs
Facultad de Psicología
Universidad Autónoma de Madrid
Campus de Cantoblanco, 28049, Madrid, Spain
E-mail: david.jacobs@uam.es
Abstract

The theory of direct learning portrays learning as specificity between higher-order informational quantities, referred to as information for learning, and change in performance that occurs with practice (Jacobs & Michaels, 2007). The focus of the theory is on the lawful generation and possible use of information for learning. The present study illustrates and further develops the theory. Participants in the study were asked to judge the mass of unseen hand-held objects. In Experiment 1, different participants received feedback on different mechanical properties of the objects, and in Experiment 2, different participants practiced with different sets of objects. The practice led to changes in performance that, in the present portrayal, show up as movements through manifolds. As predicted by the theory, these movements are specific to information for learning, the most precise description of which is obtained with difference equations. A second and more theoretical part of the article provides a tentative formalization of the theory.
An empirical illustration and formalization of the theory of direct learning:

The muscle-based perception of kinetic properties

A recent body of work has proposed and empirically tested a direct approach to learning (Jacobs, 2001; Jacobs & Michaels, 2007; Michaels, Arzamarski, Isenhower, & Jacobs, 2008). Inspired by the direct perception approach (Gibson, 1979/1986; Michaels & Carello, 1981), we argue that there is an informational basis for learning. This means that change that occurs during learning might be specific to information for learning. In the case of judgments with feedback, as in the present experiments, such information might be found in the structure of the errors that individuals make over multiple trials. In more ecologically relevant situations, information for learning should be expected to be more complex and more difficult to identify. At the same time, biological organisms can be expected to be more sensitive to information for learning in such more ecologically relevant situations. The aim of the theory is to explain how performance improves with practice and how performance is continuously adjusted with regard to possible changes in task environments and/or changes on the side of the perceiver or actor.

Up to the present, the main application of the theory has been to explain change in the informational basis of perception and action. Such change is also the main concern of the present article. Our purpose is to advance the theory and to provide empirical support for it. Experiment 1 introduces new methodology and provides data that illustrate the theory, and Experiment 2 tests predictions derived from the theory. Both experiments concern the muscle-based perception of the mass of objects that are wielded by hand (cf. Michaels et al., 2008). Since the pioneering work of Solomon and Turvey (1988), a
considerable body of research has addressed the variables that form the informational support of muscle-based perception (e.g., Kingma, Beek, & van Dieën, 2002; Shockley, Carello, & Turvey, 2004; Turvey & Carello, 1995).

Candidate informational variables include the first principal moment of inertia ($I_1$), static moment ($SM$), mass ($M$), and the third principal moment of inertia ($I_3$).

Imagine an asymmetric object that is rotated around a particular point. Such a rotation can occur around many axes, and for some axes the object will resist more against the rotation than for others. For one axis the resistance is maximal. This maximal resistance against rotation is the object’s first principal moment of inertia. For another axis the resistance against rotation is minimal. The minimal resistance is the third principal moment of inertia. One can also define a second principal moment of inertia, but due to the symmetry properties of the objects used in the present study, the first and second moments of these objects are equal. The remaining variable, static moment, can intuitively be understood as the rotational force generated by an object if the object is suspended horizontally at the point of rotation.

In the following we also need more formal definitions. For a large class of objects, $I_1$, $SM$, and $M$ can be defined as:

$$I_1 = \int \rho(s) \, \delta(s)^2 \, dV,$$

$$SM = \int \rho(s) \, \delta(s)^1 \, dV,$$

and

$$M = \int \rho(s) \, \delta(s)^0 \, dV,$$
in which $\rho(s)$ is the mass-density function and $\delta(s)$ is the distance of a point $s$ to a fixed axis of rotation.\(^1\) The functions are integrated over the volume of the objects, $V$.

As judged from our intuitive descriptions, $I_1$, $SM$, and $M$ might seem to be very different and to some extent unrelated variables. The more formal definitions, on the other hand, reveal a common underlying structure in the variables. In fact, the only difference is the power to which the distance function $\delta(s)$ is elevated. This power is zero for $M$, one for $SM$, and two for $I_1$. Note that this is always an integer value. Would it make sense to define yet other variables by elevating $\delta(s)$ to non-integer powers? Even though such non-integer variables are not encountered as frequently in physics textbooks, biological systems may be as sensitive to them as to, say, static moment and mass (cf. Michaels & Carello, 1981; Runeson, 1977). Hence, as a first step in our theory, we aim to define large classes of informational variables, including variables that are well known from disciplines such as physics as well as less well-known informational variables.

Let us formalize this again. For each real number $x$ we define the informational variable

$$I(x) = \int \rho(s) \ \delta(s)^x \ dV,$$

(1)

in which $x$ remains constant with regard to the integration. Eq. 1 can be understood as a space coordinated by $x$. Each locus in the space, or each $x$, defines an informational variable. The space includes the variables $M$, $SM$, and $I_1$ at loci 0, 1, and 2, respectively, and it includes other variables at other loci, including the non-integer numbers between 0,
1, and 2. Such spaces are referred to as information spaces (Jacobs & Michaels, 2007; Michaels et al., 2008; cf. Smith, Flach, Dittman, & Stanard, 2001).

A peculiarity of the space captured by Eq. 1 is that only one coordinate, $x$, is required to identify a variable in the space. One can also define higher-dimensional spaces. An example of a 2-dimensional space is given by the equation

$$I(x,y) = y I_3 + (1-y) \int \rho(s) \delta(s) x dV .$$

(2)

As in the 1-dimensional case, each locus in this space, or each pair $(x,y)$, represents an informational variable. The variables $M$, $SM$, and $I_J$ are now included at the loci $(0,0)$, $(1,0)$, and $(2,0)$, respectively. As an aside, the fact that some variables in such spaces (i.e., $I_3$) are labeled more easily than other variables [i.e., $0.2 I_3 + 0.8 \int \rho(s) \delta(s) 1.3 dV$] does not imply that these variables have a different status in the theory. All loci are equal in this regard.

The theory holds that, at particular moments, individuals use one of the informational variables in the space. We say that the individuals are localized at the point in the space that represents the variable that they use at that moment. With learning, observers typically show changes in variables use (Andersson & Runeson, 2008; Fajen & Devaney, 2006; Jacobs, Runeson, & Michaels, 2001; Michaels & de Vries, 1998; Wagman, Shockley, Riley, & Turvey, 2001; Withagen & Michaels, 2005). Changes in variable use are commonly referred to as attunement or as the education of attention (Michaels & Carello, 1981). In our portrayal, such changes show up as movements through information spaces. We have previously observed such movements in the
muscle-based perception of the length of objects (Michaels et al., 2008) and in other perceptual tasks (Jacobs & Michaels, 2007).

As a final preparatory step for the theory, note that movements through information spaces are dual to changes in the functional arrangement of biological tissue. Imagine, for example, an individual wielding an object. During the wielding, a large number of muscle spindles and Golgi tendon organs are continuously active in many muscles and tendons. It is not unreasonable to hypothesize that detecting a particular informational variable means (1) wielding in a particular way and (2) extracting a particular invariant pattern from the spatially distributed activity of the muscle spindles and Golgi tendon organs (Fitzpatrick, Carello, & Turvey, 1994). One might likewise hypothesize that changes in variable use, or movements through information spaces, go together with changes in the functional arrangement of the biological tissue that controls the exploratory movements and extracts or integrates the receptor activity.³

The central claim of the direct learning theory, then, in its most general form, is that learning is nothing but specificity between change in biological tissue and information for learning. As applied to the more restricted case of perceptual attunement, the claim means that movements through information spaces are specific to information for learning. Hence, direct learning does not require inferential processes on the part of an intelligent supervisor (homunculus), nor does it require hypothetical constructs (representations) as mediators between organism and environment. Instead, the direct learning approach requires a careful consideration of the situation in which learning takes place (cf. Fowler & Turvey, 1978), or, more precisely, of the way in which information
for learning is generated. Such considerations replace the hypothetical cognitive processes that are central to other approaches to learning.

Let us mention that the theory does not consider learning as a process that has a beginning and an end. Learning does not start or stop. Our work is therefore closely akin to research efforts that aim to explain how perceiving and acting are continuously updated and maintained (e.g., Bingham & Pagano, 1998; Bruggeman, Zosh, & Warren, 2007; Rieser, Pick, Ashmead, & Garing, 1995) as well as to research efforts that describe how people learn to perceive and act in novel situations (e.g., Fajen, 2008a, 2008b; Montagne, Buekers, Camachon, de Rugy, & Laurent, 2003). As illustrated by the following experiments, the present study is also closely akin to studies that apply concepts from the theory of differential equations, such as manifolds and vector fields, to the understanding of perceiving and acting (e.g., Jirsa & Kelso, 2005; Warren, 2006) and, especially so, to studies that apply these concepts to the understanding of learning (e.g., Newell, Liu, Mayer-Kress, 2001; Schöllhorn, Mayer-Kress, Newell, & Michelbrink, 2009; Shaw & Alley, 1985; Shaw, Kadar, Sim, & Repperger, 1992).

Experiment 1

The purpose of this experiment is to test the usefulness of our candidate information spaces (Eqs. 1 and 2), to demonstrate that learning entails movement through the spaces, and to propose quantities that might qualify as information for learning. Participants practiced with feedback to judge the mass of unseen tensor objects (Figure 1; Amazeen & Turvey, 1996). We chose to use tensor objects because these objects allow us to create sufficient variation in the considered informational variables. All participants were instructed to judge the mass of the objects. However, a mass group received
feedback on mass, a static moment group received feedback on static moment, and a control group did not receive feedback. The experiment extends results reported in Michaels et al. (2008), in which we studied the perception of length, to the muscle-based perception of mass. The results of the present experiment will also allow us to explore difference equations as a means to study trial-by-trial performance.

Method

Participants. Twenty-two students of the University of Connecticut were randomly assigned to three groups; 9 to the mass group, 9 to the static moment group, and 4 to the control group. The results of one participant in the mass group will not be reported because he behaved erratically throughout the experiment and his judgments did not correlate with mass in the posttest, very much unlike the judgments of other participants.

Apparatus. Eighteen tensor objects were used. The length of the central rod of the objects was 47 cm and the length of the crossbars was 52 cm. The wooden frames of the objects weighted about 172 g. Four metal weights were attached to each object. The four weights on each object were equal, but different weights were used for different objects. Likewise, the four weights on each object were attached at an equal distance from the central rod, but different distances were used for the different objects. The objects also differed in the distance between the hand-held end of the object and the place at which the crossbars were attached to the central rod. Table 1 provides more detail about the set of objects.
Procedure. Participants were seated in a chair with an armrest for the right arm. A curtain prevented them from seeing objects held in their right hand. They held a reference object in their left hand. On different trials participants were given different objects in their right hand. They were asked to wield the objects as they wished and to judge the mass of the object in their right hand as compared to the reference object. The judgments were made by moving a slider along a (continuous) vertical scale. If the objects were judged to be equal, the slider was to be placed in the middle of the scale, where the word “equal” was written. If the object in the right hand was judged to be heavier, the slider was to be placed on the upper part of the scale; the heavier the object was judged to be, the higher the slider was to be placed. The lower part of the scale was used for the lighter objects. The words “heaviest” and “lightest” were written at the upper and lower extremes of the scale. Participants could place the reference object on a table and thereby free their left hand whenever they wished. This allowed them to adjust the position of the slider with their left hand.

Design. The experiment consisted of a 17-trial pretest, three 17-trial practice blocks, and a 17-trial posttest. Object 18 was always used as reference object; in each block of trials the other 17 objects were given to the participants in their right hand in a random order. No feedback was given in the pretest and posttest. After each practice trial, participants in the mass group received feedback on mass. The experimenter provided the feedback by placing the slider at the position corresponding to the actual mass of the object. The position corresponding to the actual mass was computed by linearly
transforming the mass-values of the 17 objects so that the mass of the heaviest object corresponded to the upper extreme of the scale and the mass of the lightest object to the lower extreme. This linear transformation associated in-between values on the scale to the other objects.

Participants in the static moment group received feedback on static moment after practice trials, which means that the slider was placed at the upper extreme of the scale for the object with the highest static moment, at the lower extreme for the object with the lowest static moment, and at in-between values corresponding to their static moment for the other objects. Participants sometimes expressed suspicion with regard to the veridicality of the feedback. This happened more frequently in the mass group, in which the feedback was in fact veridical. If participants expressed suspicion we insisted that the feedback was veridical with regard to the mass of the objects and that we would prove this to them after the experiment. After the experiment we showed a few objects for which mass judgments are often inaccurate to curious and/or suspicious participants. Participants in the control group did not receive feedback. The experiment lasted about 1 hour.

Results and Discussion

In this section we describe how well participants were able to perform the task, show that learning can be portrayed as movement through information spaces, and show that such movements are specific to information for learning. We consider information for learning with block-by-block analyses and with more speculative trial-by-trial analyses. We also address the informational basis of calibration.
Performance. Table 2 presents the average correlations of the judgments with mass and static moment. The most interesting correlations are the ones between the judgments and the feedback variables of the respective groups. In the mass group, the average correlation between the judgments and mass increased from .64 in the pretest to .80 in the posttest, $t(7)=6.7, p<.01$, single-tailed. In the static moment group, the average correlation between the judgments and static moment seemed to increase from .80 in the pretest to .88 in the posttest, although this increase was only marginally significant, $t(8)=1.5, p<.10$, single-tailed. Together these results seem to confirm that performance improved with regard to the feedback variables.

We will see in the next subsection that this improvement can at least partly be attributed to a convergence on the feedback locus in the information space. Note that the correlation between static moment and the judgments of participants in the control group also improved, from .73 in the pretest to .91 in the posttest, $t(3)=4.7, p<.01$, single-tailed. The current version of the direct learning theory does not account for this decrease in variability in the absence of feedback. None of the remaining pretest-posttest comparisons of the correlations shown in Table 2 were significant or marginally significant.

Information for perception. We calculated which locus in the space defined by Eq. 2 correlated most highly with the judgments for each participant on each block of trials. These highest-correlating loci were interpreted as the informational variable that was exploited by that participant on that block. Figure 2 presents the standard error
ellipses of these loci for each group of participants and each block of trials. Standard
error ellipses indicate a region that includes about 40% of the data points. The ellipses in
Figure 2 thus indicate which informational variables participants appeared to exploit in
the different conditions and blocks of trials.

In the pretest (top panels), participants seemed to use informational variables
somewhere between mass and static moment, mostly closer to static moment than to
mass. With practice, the distribution shifted toward mass for the mass group (left panels)
and toward static moment for the static moment group (middle panels). The average
distance between the loci of participants in the mass group and the locus representing
mass decreased from .88 in the pretest to .38 in the posttest, \( t(7)=4.8, p<.01 \), single-tailed,
and the average distance between the loci of participants in the static moment group and
static moment showed a marginally significant trend toward decreasing, from .50 to .34,
\( t(8)=1.4, p<.10 \), single-tailed. In sum, with practice participants seemed to move toward
the feedback variables. The experiment did not reveal significant or marginally
significant changes in variable use for the control group.

To further test these results we performed an analysis of variance on the \( x \)-
coordinates of the loci with Group (mass, static moment, and control) as between-
subjects factor and Block (1 to 5) as within-subjects factor. Most interestingly, the Block
x Group interaction was significant, \( F(8,72)=4.5, p<.01 \), confirming the difference
between the movements through the information space. The main effect of Group was
marginally significant, \( F(2,18)=3.1, p<.07 \), indicating that the mass group tended to use
$x$-loci closer to zero, which is to say, closer to mass. Hence, we indeed observed systematic shifts in variable use, as was the case in previous studies (Michaels et al., 2008). Our next aim is to explain these shifts.

*Information for learning: block-by-block results.* Let us introduce this section with the often-cited example of gannets that dive to capture fish and fold their wings at a critical value of, say, either optical size or $\tau$ (Lee & Reddish, 1981). Among the observable consequences of using $\tau$ might be the fact that the majority of the dives are successful (i.e., without early or late wing withdrawals and the negative consequences thereof). The use of optical size, on the other hand, might lead to folding the wings too early for big fish and too late for small fish. The theory of direct learning holds that the observable consequences of using a particular variable are characteristic of (i.e., specific to) the used variable and, more importantly, that learning is based on such observable consequences. The crucial question is: Which of the many observable consequences do learners use?

To come back to our experimental situation: What are the observable consequences of basing mass judgments on a particular locus of our information space? Consider a few particular cases. First, individuals who use static moment and receive feedback on mass will overestimate tensor objects with the masses placed relatively far from axis of rotation, because such objects are high on static moment. For the same reason, they will underestimate tensor objects with the masses placed close to the rotation axis. Hence, among the consequences of using static moment to judge mass is a positive covariance between the errors that are made and the average distance of the mass of the object to the rotation axis, which is to say, a positive covariance between the errors and
the ratio of static moment and mass. Likewise, among the consequences of using mass to judge static moment is a negative covariance between the errors and the ratio of static moment and mass.

Let us formalize and generalize this. Let $\text{cov}\{E, SM/M\}$ denote the covariance over multiple trials between (1) the judgments minus the feedback, or the error, $E$, and (2) static moment divided by mass. Using Eqs. 1 or 2, one can check that if the $x$-locus of the used informational variable is larger than the $x$-locus of the feedback, the observable quantity $\text{cov}\{E, SM/M\}$ tends to be positive, and if the $x$-locus of the used variable is smaller than the one of the feedback, $\text{cov}\{E, SM/M\}$ tends to be negative. This means that $\text{cov}\{E, SM/M\}$ is informative about the cause of the errors in the current performance and thereby also indicative about how to improve performance: If the quantity is positive (negative), the learner should move to the left (right). Note that the previous reasoning concerns the $x$-axis of the space captured by Eq. 2. One can check that the quantity $\text{cov}\{E, I_3\}$ is equally informative with regard to the $y$-axis.

Of course, other informative quantities should be expected to reside in the structure of the errors that individuals make over multiple trials. According to the direct learning theory, one of these quantities must have guided the learning observed in the previous section. Much work is required to explore these quantities, and at present we have to make a considerable leap to pick out one of them. However, such a candidate would allow us to further illustrate the theory and to introduce new methodology. In the following we therefore elaborate the hypothesis that the observed movements through the information space are specific to (i.e., guided by) the above-mentioned covariances.
In short, as a first step to explain the shifts in variable use we consider the equations

\[ \dot{x}(t) = -k_1 \text{cov}\{E, SM/M\} \]  
\[ \dot{y}(t) = -k_2 \text{cov}\{E, I_3\}, \]  

in which \(E\) is the error, or the judgment minus the feedback, and \(k_1\) and \(k_2\) are positive constants. Eq. 3 holds that the momentary change in \(x\) is specific to the covariance over multiple trials between \(E\) and \(SM/M\) (look ahead to Eq. 7 for more detail). Eq. 4 holds that the momentary change in \(y\) is specific to the covariance between \(E\) and \(I_3\).

---

Insert Figure 3

---

If \(\dot{x}\) and \(\dot{y}\) are interpreted as the horizontal and vertical components of a vector at the locus \((x,y)\) of the space defined by Eq. 2, then, given particular task constraints, Eqs. 3 and 4 can be represented as a vector field on this space. Figure 3 presents these fields for the mass and static moment groups. The vector field could not be computed for the control group because no feedback was given in this group. To compute the fields we assumed that the covariances in Eqs. 3 and 4 are defined over 17 trials (the block size). We also assumed that the judgments would be proportional to the used informational variables or loci in the space. This means that the judgments and thus the predicted errors
are different for the different loci in the space, which explains why the vectors are different for different loci. The fields are different for the different groups because the feedback and thus the predicted errors were different for the groups. Figure 3 also shows the pretest and posttest ellipses for the three groups of participants. Note that the ellipses flow with the fields, which is to say, the information vectors partly explain the direction and magnitude of the changes in variable use.

To test this observation, we computed a vector indicating the change in variable use from one block of trials to the next block of trials for each block of trials of each participant. These vectors are presented in Figure 4. The base points of the vectors indicate the variable used by a particular participant on a particular block of trials. The direction of the vectors indicates the direction of the change in variable use to the next block of trials and the magnitudes of the vectors are proportional to the magnitudes of the change in variable use. To compare the vectors in Figure 4 with the vectors in Figure 3 at the same loci we use a vector correlation discussed by Crosby, Breaker, and Gemmill (1993). The square of this correlation is a generalization of the square of the Pearson product-moment correlation. In a two-dimensional case, the square of the correlation is a number between zero (no correlation) and two (perfect correlation).

This squared correlation was .85 for the mass group and .99 for the static moment group. As suggested by Crosby et al. (1993), we tested the squared correlations using a chi-square distribution with 4 degrees of freedom. In both cases the vector correlation differed significantly from zero; $\chi^2 = 27.1, p<.001$, for the mass group, and $\chi^2 = 35.5,$
Direct Learning

$p<.001$, for the static moment group. The squared correlation for the control group, computed with regard to the static moment field, was .49, which is lower than for the other groups. This correlation did not differ significantly from zero, $\chi^2 = 7.8, p>.10$. However, as also indicated by Crosby et al., the use of the chi-square statistic is less justified for the control group because for this group we had only 16 observations (4 participants x 4 blocks).

Information for learning: trial-by-trial results. We next explore the use of difference equations to obtain a more precise description of trial-by-trial performance. Difference equations are an appropriate tool because of the discrete nature of the task. In the previous subsection we conveniently defined information for learning over 17 trials (the block size). The present analyses include a speculative estimation of how many previous trials are in fact used. To address this issue, we first need to reconsider the definition of covariance. An estimate of the covariance of $v = \{ \ldots, v_{n-2}, v_{n-1}, v_n \}$ and $w = \{ \ldots, w_{n-2}, w_{n-1}, w_n \}$ based on the $m$ trials that precede trial $n+1$ is given by

$$
cov_{m,n}(v,w) = \frac{1}{m} \sum_{i=1}^{m} (v_{n-m+i} - \bar{v}_{m,n})(w_{n-m+i} - \bar{w}_{m,n}),
$$

in which $\bar{v}_{m,n}$ is the mean of $v$ from trial $n-m+1$ to trial $n$ and $\bar{w}_{m,n}$ is the mean of $w$ from trial $n-m+1$ to trial $n$. In the following analyses we use the alternative definition

$$
COV_{m,n}(v,w) = \sum_{i=1}^{m} i^*(v_{n-m+i} - \bar{v}_{m+1,n})(w_{n-m+i} - \bar{w}_{m+1,n}).
$$
Note that, in this alternative definition, \( COV \) is written in capital letters. The parameter \( c \) indicates how the different values of \( v \) and \( w \) are weighted on different trials. If \( c=0 \) all values contribute equally to the covariance, as in Eq. 5, and more recent values contribute relatively more if \( c>0 \). Another difference between Eqs. 5 and 6 is that the means in Eq. 6 are based on \( m+1 \) values rather than on \( m \) values. This allows us to meaningfully consider the covariance for \( m=1 \). Finally, because the following analyses are insensitive to multiplication by a constant, we deleted the factor \( 1/m \) from the definition.

We can now return to the purpose of this section: to understand trial-by-trial performance and estimate the number of trials that are used for learning. Approximating the horizontal change in variable use, \( \dot{x}(t) \) in Eq. 3, by the change in the horizontal location from trial to trial, \( x_{n+1}-x_n \), and using the alternative definition of covariance, we obtain that

\[
x_{n+1}-x_n = -k_1 \ COV_{m,n} \{ E, SM/M \} .
\]

The right-hand side of Eq. 7 can directly be computed from the experimental data, for each trial of each participant and for each \( m \). We next introduce a procedure to estimate the left-hand side of Eq. 7.

The purpose of the following procedure is to obtain trial-by-trial estimates of which variables are used (i.e., of the \( x_n \)'s). We limit the analyses to the 1-dimensional information space captured by Eq. 1, and we first illustrate the procedure for a single tensor object. Figure 5 gives the relation between the informational variables or loci (horizontal axis) and the judgment that observers are predicted to make for Object 10
(vertical axis). For each informational variable the predicted judgment is scaled between 50 (upper extreme of the response interval) and -50 (lower extreme of the response interval). The predicted judgment corresponding to the use of static moment $(x=1)$ is low. This is because the masses on this particular object are placed close to the handle (look back to Table 1). As a consequence, the static moment of the object is low. However, the object is in fact heavier than most other objects, which explains that the use of loci closer to zero predicts higher judgments (i.e., the curve is higher on the left).

-------------------
Insert Figure 5
-------------------

Knowing the relation between the used informational variables and the predicted judgments, or knowing the curve given in Figure 5, allows one to estimate which variable was used for a particular judgment. That is, given a particular judgment, one can tentatively conclude that the observer used the variable whose prediction best approximates the judgment. For example, as illustrated in Figure 5, a hypothetical mass judgment of 20 for Object 10 would lead one to conclude that the observer used the variable at locus $x=0.24$, and a hypothetical judgment of -15 would go together with the variable at locus $x=1.13$. A necessary condition for this determination is that the judgments predicted by the different $x$-loci must be sufficiently different. For Object 10 this was the case because the curve in Figure 5 is sufficiently steep.

We computed curves such as the one in Figure 5 for all tensor objects. The condition of sufficient steepness held for 9 of the 17 tensor objects (Objects 5-12 and Object 14; see Table 1). This means that we could estimate the $x$-loci of each individual participant on 45 trials (9 trials x 5 blocks) out of the 85 trials (17 trials x 5 blocks). We
computed the predicted x-loci for all of these trials. We then averaged these x-loci for each block of trials and for each experimental condition. The averaged loci are presented in Figure 6, which shows that the mass group (circles) seemed to converge toward mass (x-locus=0.0), whereas the static moment group (diamonds) and the control group (squares) did not. The trial-by-trial analyses therefore replicate the previously observed shifts in variable use at least partly, which means that the trial-by-trial loci might be a useful measure to estimate the number of trials used for trial-to-trial changes in variable use.

After estimating the x-loci, we computed the change in the x-locus (i.e., the left-hand side of Eq. 7) for each practice trial of each participant (with the exception of the first 10 practice trials). The changes in x-locus were correlated with the right-hand side of Eq. 7 for each value of \( m \) between 1 and 10 and using the best-fitting coefficient \( 0 \leq c \leq 5 \) for each participant (see Eq. 6). These correlations are presented in Figure 7. The changes in x-locus seemed to be best predicted with \( m=4 \). For \( m=4 \), the correlations differed significantly from zero both in the mass group, \( t(7)=4.5, p<.01 \), single-tailed, and the static moment group, \( t(8)=7.5, p<.001 \), single-tailed. Furthermore, for \( m=4 \), the difference between the correlations for the mass and control group was marginally significant, \( t(10)=1.34, p=.10 \), single-tailed, and the difference between the static moment group and the control group was significant, \( t(11)=2.0, p<.05 \), single-tailed. We conclude that our best current estimate about the number of trials that are used for learning is about 4. However, due to the speculative nature of the analyses, we prefer to draw attention to the
applied methodology rather than to this conclusion on itself. To complete the trial-by-trial analyses we address the informational basis of calibration.

-------------------
Insert Figure 7
-------------------

**Information for calibration: trial-by-trial results.** In order to study calibration with the tools explored in the previous section, we first need a more precise portrayal of perception. Perception, $P$, can be described as a single-valued function, $f$, of an informational variable, $I$, so that $P=f(I)$. In a trial-by-trial analysis this equation becomes $P_n=f_n(I_n)$, in which $n$ indicates the trial number, so that $I_n$, for example, refers to the value of the operative informational variable on trial $n$. In our experimental situation, the value of the informational variable depends on the particular tensor object used on the trial, $o_n$, and the locus, $x_n$, of the individual in the 1-dimensional space described in Eq. 1. Hence, the equation can be written as $P_n=f_n(I(x_n,o_n))$.

In the previous section we addressed change in variable use, referred to as the education of attention or attunement. In the present section we address change in the single valued function, $f$, referred to as calibration (e.g., Jacobs & Michaels, 2006). We consider the simplest case possible in which $f$ is a linear function with slope 1, so that the function for perception on trial $n$ becomes

$$P_n = I(x_n,o_n) + c_n .$$

The parameter $c_n$ is a calibration constant. The aim of this section is to identify information that specifies trial-to-trial variation in $c_n$.5
Consider the equation

\[ c_{n+1} - c_n = -k \sum_{i=1}^{m} i^i E_{n-m+i}, \]  

(9)

in which \( E_{n-m+i} \) is the error made on trial \( n-m+i \) and \( k \) is a positive constant. This equation holds that the value of \( c_n \) decreases if mass (or static moment) is overestimated on previous trials and increases if the mass (or static moment) is underestimated on previous trials. As in Eq. 6, \( m \) indicates the number of trials used to define information for calibration and \( c \) is a coefficient that indicates how heavy recent trials are weighted with respect to earlier trials.

Again, the right-hand side of Eq. 9 can be computed from the experimental data, for each \( n \) and \( m \) and for each participant. To estimate the left-hand side of the equation we first calculated which informational variable in our 1-dimensional space (Eq. 1) best predicted the judgments by determining the locus in the space that correlated most highly with the judgments per block of trials. We then took the difference between the predicted and actual judgments on a trial as an estimate for the calibration constant on that trial. Figure 8 presents the average correlation between the changes in calibration from trial to trial on practice trials (left-hand side of Eq. 9) and the information for calibration (right-hand side of Eq. 9) as defined over different numbers of trials (\( m \) in Eq. 9), using the coefficient \( c \) that led to the highest correlation for each \( m \) and each participant.

-------------------

Insert Figure 8
-------------------
Figure 8 illustrates that the best predictions were obtained by defining the information for calibration over a single trial ($m=1$), both in the mass group and in the static moment group. For $m=1$, the predictions were significantly better in the mass group and static moment group (continuous lines) than in the control group (dashed lines); $t(10)=1.87$, $p<.05$, single-tailed, and $t(11)=4.77$, $p<.001$, single-tailed, respectively. Even though the superiority of the predictions in the mass group and static moment group over the control group is significant, the fact that substantial correlations are found also in the control group indicates that these correlations cannot entirely be attributed to calibration mechanisms as we have defined them.

To further test the conclusion that information for calibration should be defined over a single trial we computed the correlations between our predicted calibration constant on trial $n+1$ (for practice trials) and the error observed on trial $n$, on trial $n-1$, on trial $n-2$, and on trial $n-3$. These correlations are presented in Figure 9. The calibration constant on trial $n+1$ correlates negatively with the error observed on trial $n$ both in the mass group and in the static moment group. We interpret this negative correlation as evidence for a calibration mechanism (specificity) that adjusts the calibration constant on the basis of the error on the previous trial. For trial $n$, the difference between the mass group and the control group and between the static moment and the control group are both significant, $t(10)=3.9$, $p<.01$, single-tailed, and $t(11)=2.7$, $p<.05$, single-tailed, respectively. No significant differences were observed for trials $n-1$, $n-2$, and $n-3$. 

---------------------
Insert Figure 9
---------------------
Let us summarize the trial-by-trial portrayal. Perception is described by Eq. 8. This equation stands for the coupling, or specificity, between an ambient energy property (in our case an inertia-like property) and a psychological phenomenon (the perception of mass). The specificity is established by a smart perceptual device (Runeson, 1977) that is hypothesized to integrate (detect) information spread over receptor arrays. Learning and calibration are described by Eqs. 7 and 9. These equations describe the specificity between time-wise extended ambient energy properties (the right-hand sides of the equations) and change in the functional arrangement of biological tissue (i.e., in the smart perceptual device). This change, in turn, is portrayed as movement through information and calibration spaces (the left-hand sides of the equations).

Experiment 2

In the preceding we have argued that learning goes together with movements through information spaces (cf. Jacobs & Michaels, 2007; Michaels et al., 2008). Information that specifies such movements is exemplified by Eqs. 3 and 4. As illustrated in Figure 3, these equations can be represented by vector fields. The present experiment is based on the observation that, given the equations, different sets of objects go together with different vector fields, and thereby make different predictions about the course of learning. The experiment tests such predictions. Because the predictions are derived from the theory of direct learning, the experiment can be interpreted as a test of the theory.

We created two practice sets of tensor objects. The sets differed in the combinations of the weights and the locations of the weights on the objects and, as a consequence, in the vector fields associated with Eqs. 3 and 4. For the first set, the fast set, the vectors were long, predicting a fast movement through the information space. For
the second set, the *slow set*, the vectors were short, predicting a slower movement. In both sets, for 8 objects the weights were placed close to the axis of rotation and for 8 objects the weights were placed at a position farther from that axis. For the fast set, heavier weights were more often placed at the nearer location and lighter weights more often at the farther location, and vice versa for the slow set. This indeed resulted in longer and shorter vectors, respectively (look ahead to Figure 10).

Why is this the case? Imagine an individual who judges mass on the basis of a variable near static moment. The individual will overestimate objects with the masses placed far from the axis of rotation, because such objects are high on static moment as compared to mass, whereas he or she will underestimate objects with the masses placed near the rotation axis. Hence, participants in the fast group are predicted to overestimate the mass of many light objects and underestimate the mass of many heavy objects, leading to a large covariance in Eq. 3 and, hence, to long information vectors. Participants in the slow group, in contrast, are predicted to judge many heavy objects as being more heavy and many light objects as being more light, which does not result in the same pattern of errors on the fixed response scale and therefore leads to shorter information vectors.

Remember that a few participants in Experiment 1 were suspicious about the veridicality of the feedback, especially in the mass group, for which the feedback was in fact veridical. These observers might have been suspicious because of the relatively large distance between the initially used loci and the locus of mass. As in Experiment 1, participants in the present experiment were asked to judge mass, but they received feedback based on the variable $(.2,0)$ of the space described by Eq. 2. This feedback
variable is slightly closer to the variables that participants tend to use initially than the feedback variable \((0,0)\), or mass, which might reduce the participants’ suspicion with regard to the feedback.

**Method**

Experiment 2 was identical to Experiment 1 with the following exceptions. Nine students of the University of Connecticut were assigned to the fast group and 9 to the slow group. For both groups, feedback was given by placing the slider at the position corresponding to the object’s value on the variable \((.2,0)\) in the space captured by Eq. 2. The difference between the groups was the placement of the weights on the tensor objects in the 3 practice blocks. The fast group practiced with the fast set and the slow group practiced with the slow set (see Table 3). The test-phase set was used in the pretests and posttests of both groups. Note from Table 3 that Objects 9 to 17 of the different sets were identical. Also, Objects 1 to 8 of the test-phase set were identical to Objects 9 to 16. This allowed us not to change the weights between the test phases and practice; we simply used Objects 9 to 16 twice in the test phases.

```
Insert Table 3
```

**Results and Discussion**

The average correlation between the judgments of participants in the fast group and the feedback variable was .82 in the pretest and .85 in the posttest. These correlations were .84 and .83 for the slow group. No significant effects were observed in an analysis of variance on the \(z\)-scores of the correlations; \(F(1,32)=0, p=.98\), for differences between the groups, \(F(1,32)=.13, p=.72\), for the change with practice, and \(F(1,32)=.6, p=.44\), for
the interaction. From our hypothesis one might have expected the improvement to be larger in the fast group. This was not found to be the case. Even so, for a more direct test of the hypothesis we next compare the movement through the information space for both groups.

Figure 10 presents the standard error ellipses for the two groups, in pretest and posttest, along with the vector fields of Eqs. 3 and 4 for the corresponding sets of tensor objects. The average distance between the loci of participants in the fast group and the feedback variable was .71 in the pretest and .44 in the posttest; $t(8)=1.97$, $p<.05$, single-tailed. For the slow group the average distance was .56 in the pretest and .57 in the posttest; $t(8)=.10$, $p>.5$, single-tailed. This is consistent with our hypothesis: The fast group changed quickly in variable use whereas the slow group did not seem to change in variable use. The information-based (or direct) account of learning attributes this apparently faster change to the fact that the candidate information for learning specified a faster change (i.e., had longer information vectors in Figure 10).

An alternative interpretation holds that movements through information spaces are gradient-based rather than information-based. We refer to a movement with the term gradient-based if participants detect how much their performance would improve or deteriorate if moving in a particular direction and subsequently move in the direction that leads to the largest improvement. Consider Figure 11. The horizontal planes in the graphs represent the two-dimensional information space captured by Eq. 2. The curved surfaces in the upper and middle graphs are the squared correlations between the informational
variables (i.e., loci) and the feedback variable, for the sets of objects used in the fast and slow groups, respectively. The surface is steeper and lower for the fast group than for the slow group, meaning that a gradient-based account can also explain our results.\(^6\)

As an aside, the lower graph in Figure 11 illustrates the results of the fast group with Gaussian probability density functions. Such a representation might in some cases be a useful alternative to the standard error ellipses used in the present article (cf. Jacobs & Travieso, 2009).

To summarize, the experiments reported in the preceding provide an application of the direct or information-based approach to learning to the muscle-based perception of mass. The aim of the experiments was to illustrate and advance the theory in the context of this particular example. The following sections go beyond examples and tentatively aim to advance the theory more generally, abstracted from particular embodiments in physical and biological tissue. We try to precisely define the concepts that are involved in the theory and to make them as consistent as possible with basic mathematical tools. It is our belief that the to-be-presented formalization is worthwhile. Nevertheless, readers not interested in more formal detail might also want to proceed to the general discussion immediately.

A classification of direct learning tasks

Classifications are among the most elegant achievements of science. In mathematics, for instance, classifications are ubiquitous and of utmost importance. Let us address a simple example. If a vector field associates a zero vector to a point \( x \in \mathbb{R}^2 \), then
\( x \) is called a critical point of the field. With several additional assumptions, critical points can be classified as (1) stable nodes, (2) unstable nodes, (3) saddle points, (4) centers, (5) stable spirals, and (6) unstable spirals. This particular classification holds that, up to isomorphism, there are only six possible vector fields in neighborhoods of critical points. The fact that only one type of stable nodes is included, for instance, means that any two stable nodes in \( \mathbb{R}^2 \) are considered to be isomorphic.\(^7\)

Remember that giving a vector field is equivalent to giving a system of differential equations. The curves that follow the vector field are the solutions of the system of equations. These curves are referred to as the flow curves of the field. Figure 12 presents examples of vector fields and flow curves for the stable node and the stable spiral. Both are referred to as point attractors. The spiral has the property that, however close one gets to the attractor point, the flow curves will always show an infinite number of turns around the attractor point before getting there. In a stable node, in contrast, the flow curves can only show a finite number of turns around the attractor point.

\[\text{-------------------}\]
\[\text{Insert Figure 12}\]
\[\text{-------------------}\]

In mathematics this difference is essential. In many cases one cannot transform a stable spiral into a stable node precisely because one cannot transform an infinite number of turns into a finite number of turns. For our theory the difference is not essential. In systems with noise it does not make sense to consider an infinite number of turns in arbitrarily small neighborhoods. We therefore assume that systems cannot approach point attractors with an infinite number of turns. This exclusion of spirals from the analysis means that, up to mere ways of describing the system, or up to isomorphism, there is only
one two-dimensional point attractor—the stable node. We likewise assume point
attractors in higher-dimensional systems to be isomorphic.

In the following we describe how this assumption allows us to obtain tentative
classifications of learning tasks and systems. However, even though the arguments are
structured around them, our purpose is not only to obtain the classifications. We are as
much or perhaps even more interested in the definitions that are needed on our way. This
includes definitions concerning the ecological view on perceiving and acting (Gibson,
1979/1986) as well as definitions concerning direct learning (Jacobs & Michaels, 2007).

**Definition 1.** A *direct learning task* is a task in which perceivers, if involved in the
task, intend to perceive a property during a substantial amount of time, without changes
in the intention.

Examples of direct learning tasks include judging a particular property on
successive trials in an experiment or continuously perceiving the distance to a car driving
ahead of one’s own car. Direct learning tasks also include actions such as catching or
hitting balls on successive occasions, or maintaining postural balance, even though such
actions might require minor changes in terminology. Note that we have defined direct
learning rather than direct perception tasks. Usually, however, perceiving implies
learning, which means that similar definitions can be used for direct perception and direct
learning tasks. Also note that Definition 1 excludes changes in intention, which are not
considered in the present manuscript. We refer the reader to Arzamarski et al. (2009) for
changes in intention as interpreted from a direct-learning perspective.

**Definition 2.** A property that perceivers intend to perceive in a direct learning task
is referred to as a *to-be-perceived property*. 
In our experiments the to-be-perceived property is mass, and in the previously mentioned driving task the to-be-perceived property is the distance to the car that is driving ahead. In general, however, one should expect to-be-perceived properties to be the more apparently complex properties that are referred to as affordances (Gibson, 1979/1986; Turvey, 1992).

**Definition 3.** Detectable ambient energy variables that specify to-be-perceived properties are referred to as *Gibsonian information*.

Under certain boundary conditions (e.g., a constant object size), *tau* might qualify as Gibsonian information about the time-to-contact of approaching objects (Lee & Reddish, 1981). Likewise, the focus of optical expansion constitutes information about heading direction, and texture gradients about the slant of surfaces (Gibson, 1979/1986). Gibsonian information is sometimes simply referred to as *information for perception*.

**Axiom 1.** Gibsonian information exists.

This axiom means that to-be-perceived properties are typically specified by detectable ambient energy variables. The axiom restricts the theory to ecologically relevant task environments, because such environments are the ones that are sufficiently rich in information to make the axiom reasonable. The axiom is very important for ecological psychology; one might even argue that the discipline of ecological psychology is a consequence of the axiom. That is, assuming the existence of information has often led to subsequent efforts to identify such information. Below we extend Axiom 1 to the more encompassing Axiom 2, which concerns the existence of information for learning.
To be able to precisely formulate the latter axiom, we first need the notions of information space and variable space (Jacobs & Michaels, 2007).

**Definition 4a.** A variable space is a differentiable manifold\(^8\) each point of which represents a detectable ambient energy variable.

Eqs. 1 and 2 are examples of variable spaces. Also, if Variable A and Variable B are ambient energy variables, then \(x \text{Variable A} + (1-x) \text{Variable B}\), with \(x \in R\), is a variable space (Jacobs & Michaels, 2007). Note that variable spaces merely refer to ambient energy variables. In contrast, the more restricted notion of information space also refers to the task at hand and to the perceptual systems that perform the task.

**Definition 4b.** An information space is a variable space that includes (1) a point that represents Gibsonian information and (2) all the variables that perceivers and actors tend to use for the task. Furthermore, among those spaces, an information space is a space of minimal dimension.

Including Gibsonian information in information spaces is required for direct learning and direct realism to be congruent, because it allows the assumption that people converge on the use of Gibsonian information on the long run (Jacobs & Michaels, 2002). The term information space should be used more carefully than the term variable space.

One can easily define a variable space; the question is whether a variable space qualifies as information space. This is reminiscent of the typically careful use of the term information in ecological psychology: Ambient energy variables are information only if they satisfy the conditions of Definition 3.

**Definition 5.** A calibration space is a differentiable manifold each point of which represents a particular calibration, with the condition that it includes (1) an optimal
calibration and (2) all the calibrations that perceivers and actors use in the task. Again, among those spaces, a calibration space is a space of minimal dimension.

To give an example, if the perception, $P$, of an environmental property that an individual intends to perceive, $E$, is a linear function of a particular informational variable, $I$, say $P = aI + b$, with $a$ and $b$ varying over time and among individuals, then $a$ and $b$ are the coordinates of a two-dimensional calibration space. In this example, the use of an optimal informational variable would be evidenced by a high correlation between $P$ and $E$. An optimal calibration would be evidenced by other properties of the relation between $P$ and $E$, for instance by a low absolute difference.

**Definition 6.** A detectable ambient energy variable is referred to as information for learning if it specifies a change in variable use that eventually leads to the use of Gibsonian information.

Reconsider, for example, Eqs. 3 and 4 (cf. Michaels et al., 2008). As illustrated by these equations, information for learning should be expected to be extended over time and related to the outcome of perceiving and acting in the task environment. The information is generated by the interaction of an individual and the environment. As such, the information exists as ambient energy variable only at the moment at which tasks are actually performed. In our view, discovering the nature of information for learning in different learning tasks is one of the future challenges of the direct learning approach.

**Definition 7.** A detectable ambient energy variable is referred to as information for calibration if it specifies a change in calibration that eventually leads to an optimal calibration.
An example of information for calibration is given in Eq. 9. To simplify our definitions, we sometimes use the term *convergence processes* to refer to the education of attention and calibration (Jacobs, 2001). Information for learning and calibration is then jointly referred to as *convergence information*.

*Axiom 2a.* Convergence information exists.

Axioms 1 and 2a are in many ways similar. This similarity reflects the similarity of the theories of direct perception and direct learning. For example, the direct perception approach is a law-based approach that contrasts with more traditional symbolic, rule-based, or probabilistic approaches (e.g., Kirlik, 2009); its proponents search for laws that explain how information for perception and action is lawfully generated and laws that explain how this information guides perception and action (Turvey & Carello, 1986). Analogously, direct learning is a law-based approach to learning, the laws of which concern the generation and use of information for learning.

To continue our formalization we need a more precise interpretation of Definition 6. Given that loci in information spaces represent informational variables, vectors in information spaces represent changes in variable use, with the direction of a vector indicating the direction of the change and the magnitude of the vector indicating the speed of the change. Definition 6 therefore implies that convergence information can be represented by a vector field. This brings us to the following definition.

*Definition 8.* An *information field* is a vector field that represents information for learning.

Information fields in calibration spaces are defined likewise. The fields in Figures 3 and 10 are examples of information fields. Other examples of information fields can be
found in Jacobs and Michaels (2007) and in Michaels et al. (2008). Definitions 6 and 7 hold that information fields lead to Gibsonian information and optimal calibration, which means that the points representing Gibsonian information and optimal calibration are point attractors in information fields. This allows us to use the mathematical notion of isomorphism to define a psychological notion of isomorphism.

**Definition 9.** Direct learning tasks are *isomorphic* if they allow information and calibration spaces and information fields on these spaces that are equal up to isomorphism (1) in neighborhoods of Gibsonian information and (2) in neighborhoods of optimal calibration.

Note that this notion of isomorphism is independent of particular implementations in physical and biological tissue. For example, muscle-based perception tasks can be isomorphic to other muscle-based perception tasks as well as to auditory, visual, or audiovisual tasks. Also note that the notion of isomorphism creates groups of learning tasks, with tasks that are isomorphic to each other belonging to the same group. Let us say slightly more about the conditions under which tasks belong to the same group of isomorphism.

**Classification 1a.** Direct learning tasks are isomorphic if and only if (1) their information spaces have the same dimension and (2) their calibration spaces have the same dimension.

This classification is an immediate consequence of the fact that we consider point attractors to be isomorphic if and only if they reside in spaces of the same dimension. The classification means that, under our assumptions, direct learning tasks can be classified by two natural numbers: one for the dimension of the education of attention and one for...
the dimension of calibration. If two tasks have the same two numbers, then they are identical at our level of description, which is to say, identical with the exception of their embodiment in physical and biological structure.

Direct learning systems

The previous section concerned direct learning tasks. In that section we provided definitions related to environmental properties and informational variables. Those definitions were presented as independently as possible from observers, or perceptual systems, that might or might not perform the tasks. The present section concerns the observers’ side of the learning phenomena. In this case the definitions will be presented as independently as possible from the environmental part of the theory. Let us mention that we chose to present separate descriptions of the observers’ and environmental sides because such a separation (1) allows us to sharpen the definitions and (2) paves the way for the claim, made explicit in a subsequent section, that direct learning is, in fact, a theory about organism-environment duality. The reader will notice that the present section concerns the same phenomena as the previous section, although from a different point of view.

Definition 10. A direct learning system is a system that intends to perceive a particular property during a substantial amount of time and without changes in the intention.

Examples of direct learning systems include, say, participants in experiments, predators pursuing prey, baseball outfielders, and, perhaps with minor changes in terminology, automatic pilots or adaptive robots.
Definition 11. The functional arrangement of biological tissue that permits a direct learning system to detect (or to resonate to) a particular ambient energy variable, or have a particular calibration, is referred to as the set-up of the system.

We have speculated in the introduction that detecting a particular (mechanical) informational variable of a hand-held object might be related to (1) wielding in a particular way and (2) integrating or extracting a particular pattern of activity spread over large numbers of receptors, including muscle spindles and Golgi tendon organs (Fitzpatrick et al., 1994). The functional arrangement of the tissue that controls the exploratory wielding and extracts the invariant patterns of activity, together with the receptors themselves, might be an example of the set-up of a perceptual system. A set-up is assumed to exist (and change continuously) as long as the intention to perceive a property is maintained; if the intention changes, the biological tissue might more abruptly rearrange itself so as to become a different set-up (Arzamarski et al., 2009). Given that a particular set-up entails the detection of a particular ambient energy variable and a particular calibration (Jacobs & Michaels, 2007), the set-up, at a given moment, can be represented by a point in an information space and a point in a calibration space.

Definition 12. Subspaces of information and calibration spaces that represent all the possible set-ups of a particular direct learning system are referred to as convergence spaces.

Note that this definition on itself does not exclude the possibility that different systems might have different convergence spaces, even if they perform the same task. It might be the case, for instance, that some systems are able to detect all of the variables in an information space, meaning that their convergence space is the entire information
space. Other systems might not be able to detect all of the variables, which means that their convergence spaces are proper subspaces of the information space. Let us anticipate, however, that such situations will be excluded by Conjecture 3 in the following section.

Definition 13. A convergence path is a path in a convergence space that includes all the set-ups that occur in a particular system while the system performs a task during a substantial amount of time.

The particular convergence path that is followed through a convergence space depends on the point in the space (i.e., the set-up of the system) at the moment that one starts to observe the system, as well as on the particular system that is being observed. To give an example, for a system that does not learn, the path is a single point, possibly the point that represents the detection of Gibsonian information. (Again, however, the reader might want to look ahead to the following section, in particular to Corollary 2.)

Definition 14. A goal state is the point in a convergence space that is associated to the detection of Gibsonian information or optimal calibration.

Axiom 2b. Convergence paths (1) are smooth, (2) eventually lead to goal states, and (3) do not cross each other if they are traced by similar perceptual systems.

This axiom is meant to reflect the conditions that are required for convergence paths to be the solutions of a system of differential equations, or, equivalently, for the paths to be the flow curves of a vector field. For instance, if two convergence paths would cross each other, then there would be an intersection point from which the paths continue in different directions. This would mean that the paths cannot be described with a single vector field. Under the conditions of Axiom 2b, on the other hand, the
convergence paths are the flow curves of some vector field. This vector field can be obtained by differentiating the convergence paths.

*Definition 15.* A *convergence field* is a vector field obtained by differentiating the convergence paths of similar direct learning systems.

Axiom 2b implies that convergence paths lead to goal states, and hence that goal states are point attractors of convergence fields. This allows us to define a notion of isomorphism for direct learning systems similar to the one that was previously defined for direct learning tasks.

*Definition 16.* Direct learning systems are *isomorphic* if they allow convergence spaces and convergence fields that are equal up to isomorphism in the neighborhoods of goal states, both with respect to the education of attention and with respect to calibration.

Intuitively speaking, systems are isomorphic if, given a particular intention, they have the same flexibility in their set-ups, allowing them the same flexibility in variable use and calibration. For example, a rigid system whose convergence spaces consist of single points cannot be isomorphic to a more flexible system with higher dimensional convergence spaces, because point attractors in spaces with different dimensions are not isomorphic. The following classification states more precisely which direct learning systems belong to the same equivalence class of isomorphism.

*Classification 1b.* Direct learning systems are isomorphic if and only if (1) their convergence spaces with respect to the education of attention have the same dimension and (2) their convergence spaces with respect to calibration have the same dimension.

Again, two natural numbers are needed to classify a direct learning system, one for its dimensionality with regard to the education of attention and one for its
dimensionality with regard to calibration. Observe that the classification is not concerned with the performed task. For example, baseball outfielders might be isomorphic to other outfielders as well as to participants in perceptual experiments or automatic pilots. The classification is also indifferent with regard to factors such as, say, the curvature of the trajectories on which goal states are approached. Such factors belong to analyses at the level of particular instances of tasks and systems.

On the duality of tasks and systems

The theory of direct learning aims to study learning phenomena in natural environments, among other reasons because information for learning exists only by virtue of interactions of individuals with their environments. Even so, the previous sections separately considered concepts related to individuals and environments. The present section connects these concepts. To achieve this we explicitly formulate three conjectures. We propose that these conjectures should be accepted.

*Conjecture 1 (main conjecture).* Change in set-up is specific to convergence information.

This conjecture means that direct learning phenomena are fully explained by convergence information. The conjecture thereby sets the approach to learning apart from inference-based and knowledge-based approaches, because change that is specific to information need not be decided upon by conscious or unconscious inferential processes. Likewise, an explanatory role of internal representations is inconsistent with the conjecture because the stated specificity does not leave any room for additional explanations. Conjecture 1 leads to the following corollary.

*Corollary 1.* Axiom 2a implies Axiom 2b.
Assume that Axiom 2a is true. This implies the existence of information for learning and hence the existence of information fields. Conjecture 1, then, holds that convergence paths are specified by the vectors in an information field, which means that the convergence paths can be obtained by integrating the information field. Convergence paths obtained in this way necessarily verify the conditions of Axiom 2b. Hence, Axiom 2a implies Axiom 2b. Note that Corollary 1 would not necessarily be true in the absence of Conjecture 1. After all, properties of convergence paths would be unrelated to the possible existence of information for learning if, rather than on such information, the paths would be based on, say, inferential processing. We now consider a second conjecture.

Conjecture 2 (homogeneity of convergence information). Information fields depend on tasks rather than on systems.

This is a first conjecture about individual differences. The conjecture holds that knowing the task to be performed, including the local task constraints, allows one to also know the information field, implying that different systems (e.g., individuals) use the same convergence information. In sum, whereas Conjecture 1 holds that all individuals use information for learning, Conjecture 2 holds that they all use the same information for learning. Conjecture 2 implies the following corollary.

Corollary 2. Single points determine convergence paths.

This is because, given a convergence space and a point in the space, a path can be predicted with the information field implied by Conjecture 2 and the specificity implied by Conjecture 1. This brings us to our final conjecture.
Conjecture 3 (homogeneity of set-ups). Convergence spaces depend on tasks rather than on systems.

This conjecture holds that the set-ups of different systems (individuals) that perform the same task are equally flexible with regard to which variables they might use for the task. Conjecture 3 is not implied by Conjectures 1 and 2. To see this, imagine (1) that an information field in a two-dimensional information space specifies the convergence paths of two learning systems, (2) that one of the systems has the ability to use all of the variables in the information space, and (3) that the other system is restricted to a one-dimensional subspace. This situation is in agreement with Conjectures 1 and 2, but not with Conjecture 3. As anticipated in the introduction of this section, the conjectures connect the observers’ and the environmental sides of the learning process. This is formalized in the following corollary.

Corollary 3. Classifications 1a and 1b are dual.

This corollary means that Classifications 1a and 1b are in fact different descriptions of a single phenomenon, which is to say that tasks belong to the same group of tasks if and only if systems that perform the tasks belong to the same group of systems. Let us try to convince ourselves that this corollary is indeed implied. First, Conjecture 3 together with Definition 4b (information spaces) and Definition 12 (convergence spaces) mean that convergence spaces with respect to the education of attention qualify as information spaces. Hence, convergence spaces with respect to the education of attention have the same dimension as information spaces. Likewise, convergence spaces with respect to calibration have the same dimension as calibration spaces. It follows that all direct learning systems for a particular task are isomorphic, and that they are isomorphic
with systems for another task if and only if the tasks are isomorphic. In other words, it follows that Classifications 1a and 1b are dual.

Weakly isomorphic tasks and systems

In the previous sections we have formalized key concepts of the theory of direct learning (Jacobs & Michaels, 2007). Our aim was to relate the theory to basic mathematical tools. We also presented classifications and explicitly formulated three of the conjectures that underlie the theory. The following sections explore novel directions in the theory, starting with a reassessment of the relation between the education of attention and calibration. At certain levels of description, the education of attention and calibration seem to be clearly different. That is, change in which of the available informational variables is operative is not the same as change in the function that carries the operative variable into perception or action. At the same time, both processes are assumed to consist of specificity between convergence information and change in the set-up of perceptual systems. Why, then, should a theory distinguish the processes? We now address such concerns.

Definition 17. A complete information space is a differential manifold each point of which represents a detectable ambient energy variable together with a particular calibration, with the condition that it includes (1) a point that represents Gibsonian information together with an optimal calibration and (2) all the information/calibration pairs that observers tend to use for the task. Furthermore, among these spaces, a complete information space is a space of minimal dimension.

To give an example, complete information spaces might be obtained by taking the product of information and calibration spaces. The product of two spaces can intuitively
be understood as a higher-dimensional space that includes the same information as the spaces that form the product. Hence, a point in the product space of information and calibration spaces indeed defines a particular informational variable together with a particular calibration.

Definition 18. A complete information field is a vector field on a complete information space with the condition that the field represents a detectable ambient energy variable and that the flow curves of the field eventually converge on the point that represents the use of Gibsonian information with an optimal calibration.

Again, a complete information field could be the product of an information field with regard to the education of attention and an information field with regard to calibration. This illustrates that complete information fields describe change in informational variables as well as change in calibration. Definitions 17 and 18 allow the following notion of isomorphism.

Definition 19. Direct learning tasks are weakly isomorphic if they allow complete information spaces and fields that are equal up to isomorphism in neighborhoods (in the respective complete spaces) of the points that represent combinations of Gibsonian information and optimal calibration.

Definition 20. The dimension of a direct learning task is the dimension of the associated complete information space.

Classification 2. Direct learning tasks are weakly isomorphic if and only if they have the same dimension.

In Classification 1, the education of attention and calibration were carefully distinguished. The description offered in Classification 2, on the other hand, considers the
education of attention and calibration to be essentially equal, leading to larger and less numerous classes of equivalence. Whereas Classification 1 requires two natural numbers, one for the education of attention and one for calibration, Classification 2 requires only one natural number, namely, the number that indicates the dimension of the task.\textsuperscript{10}

Intrinsic information spaces

The present section is based on the observation that, given a particular task, many spaces might qualify as information space. For example, by stretching an information space, one obtains another information space. Below we introduce the notion of intrinsic spaces for spaces that provide parsimonious descriptions of learning phenomena. This notion is reminiscent to the notion of intrinsic information for perception (Warren, 1984). If one uses intrinsic measures (i.e., units of leg length), many perceptual systems show identical behavior, despite apparent differences that might show if one uses extrinsic measures (i.e., meters). Likewise, intrinsic spaces reveal similarities in learning phenomena for large classes of direct learning systems. These classes are the ones defined by our classifications.

Definition 21. We refer to an information field as a canonical information field if (1) it leads to Gibsonian information along the paths of geodesics and (2) the lengths of the vectors are proportional to the distance of their base points to Gibsonian information (look back to Footnote 8).

Consider, for example, the information space $\mathbb{R}^2$ with the point $(1,0)$ representing Gibsonian information. Also consider the field that associates, say, the vector $(0,0)$ to the point $(1,0)$, the vector $(-1,0)$ to the point $(2,0)$, or, defined more generally, the field that that associates the vector $(1-x,-y)$ to the point $(x,y)$. This field leads the learner to the use
of Gibsonian information along straight lines, which are the geodesics of this space, and with a speed as defined in the second part of Definition 21. The field hence is a canonical information field. Note that the precise form of an information field (or convergence field) depends on the information space that is used to study the system. That is, whereas for some spaces the convergence paths follow geodesics, for other spaces they do not. In other words, whether information or convergence fields are canonical or not depends on the used spaces. This reasoning is captured by the following definition.

**Definition 22.** Information spaces are *intrinsic spaces* if they lead to information fields that are locally canonical at points representing Gibsonian information.

Intrinsic calibration and convergence spaces can be defined likewise. Intrinsic spaces and fields are interesting among other reasons because they provide an optimal match between one’s description of the change in a direct learning system and the way in which such change is actually established by the system. One might therefore speculate that searching for intrinsic spaces might be informative about how changes in set-ups are established. To illustrate this we reconsider the example that follows Definition 21. If $x$ and $y$ are independently adjusted toward their optimal values, then a mere rescaling of the axes would result in a canonical field; the space with the rescaled coordinate parameters $x$ and $y$ would be intrinsic. If, however, the parameters $r = \sqrt{x^2 + y^2}$ and $\varphi = \arctan(y/x)$ are adjusted independently, then the convergence paths would not follow the geodesics of the space coordinated by $x$ and $y$, not even after rescaling. Rather, a space coordinated by $r$ and $\varphi$ would be intrinsic.

**General Discussion**
The present study aimed to illustrate and advance the theory of direct learning. Participants in two experiments practiced with feedback to judge the mass of unseen hand-held objects. Experiment 1 showed that, with practice, observers change in variable use and come to rely on the more useful informational variables. The changes in variable use were portrayed as movements through information spaces, and the movements themselves were argued to be specific to information for learning. These results are consistent with findings reported in Jacobs and Michaels (2007) and Michaels et al. (2008). Experiment 2 addressed predictions derived from the theory. As predicted, observers appeared to change more quickly in variable use when the information for learning specified a quicker change in variables use, and they seemed to change more slowly in variable use when the information for learning so specified. In the remaining discussion we elaborate a few key aspects of the theory.

*The structure of information spaces*

We have tentatively defined information spaces as differentiable manifolds (e.g., Euclidean spaces, surfaces in Euclidean spaces, etc.). Given that differentiable manifolds are topologies, this allows the use of topological concepts such as convergence, and given that differentiable manifolds allow a differential calculus, one can also take advantage of the theory of ordinary differential equations. At some places we even relied on concepts such as geodesics, which assumes that a metric has been chosen on the spaces. This rich mathematical structure of information spaces contrasts with previous studies that implicitly assumed information for perception and action to consist of mere sets of variables. Among such previous studies are studies about the informational basis of muscle-based perception (Kingma, van de Langenberg, & Beek, 2004; van de
Langenberg, Kingma, & Beek, 2006; Wagman et al., 2001) and some of our own studies about the visual perception of kinetic properties (Hajnal, Grocki, Jacobs, Zaal, & Michaels, 2006; Jacobs, Michaels, & Runeson, 2000; cf. Runeson, Juslin, & Olsson, 2000). We now think that portraying the available information as mere sets of informational variables is unfortunate, because more powerful learning theories can be obtained if one endows information spaces with a richer mathematical structure (cf. Smith et al., 2001).

* Toward universal equations of learning? *

Imagine a judgment-feedback task, as the one used in the present experiments, and \( n+1 \) sufficiently different informational variables, \( V_1, \ldots, V_{n+1} \). The \( n \)-dimensional space

\[
V(x_1, \ldots, x_n) = x_1 V_1 + \ldots + x_n V_n + (1-x_1-\ldots-x_n) V_{n+1}, \tag{10}
\]

with \( (x_1, \ldots, x_n) \in \mathbb{R}^n \), is a variable space for such a task. Furthermore, if change through the space is described by the system

\[
\dot{x}_i = -k_i \text{cov}(E, V_i), \tag{11}
\]

with \( i=1,\ldots,n \) and with the \( k_i \)'s being positive constants, or by more precise versions of such equations (e.g., Eq. 7), one typically observes convergence toward the more useful informational variables (cf. Jacobs & Michaels, 2007; Michaels et al., 2008).
We included Eqs. 10 and 11 in this discussion because we want to mention three critical observations related to them. First, even though the equations might describe direct learning in tasks with a judgment-feedback structure, direct learning is hypothesized to occur most particularly in more natural situations, which may lack such an explicit feedback structure. Second, Eqs. 10 and 11 are arbitrary in the sense that they give just one of the many possible systems of equations that can be used to describe information spaces and convergence. This second observation gives rise to several questions. For instance, are there mathematical theorems and proofs related to the optimality of the convergence defined by certain systems of equations? If so, do these theorems inform us about, say, how the convergence depends on statistical properties of the input variables and/or on the numbers of trials that are used in the equations? Or, to give one more example, how should one include calibration in the equations? A third observation related to Eqs. 10 and 11 is addressed in the following section.

**General-purpose versus special-purpose devices**

A general-purpose device is a device that is able to function in many different situations. A special-purpose device, in contrast, is a device that functions properly only in a more limited range of situations, because its functioning relies on particularities of the task situation (Fowler & Turvey, 1978). Eqs. 10 and 11 might inadvertently give the impression that we consider learners to be general-purpose devices. That is, the equations might suggest that, from one learning situation to the next, the learner changes the initial variables ($V_1$ to $V_{n+1}$), and then applies the same set of rules (Eqs. 10 and 11), with the result that he or she converges toward or computes the more useful information variables. In our portrayal, however, learners are hypothesized to be special-purpose devices,
meaning that evolution prepared individuals only for particular learning tasks (Fowler & Turvey). To give an example, we would speculate that perceptual systems are malleable only to the extent that this allows the systems to adapt to local ecologies that members of the considered species can be expected to occupy. In such a local ecology, a particular to-be-perceived property might be expected to be specified by a particular type of informational variables but not by others. These potentially useful variables are the ones that we expect perceptual systems to be able to attune to, and they are what we aim to describe with the notion of information space.

**Smart learning devices**

The distinction between general-purpose and special-purpose devices is related to Runeson’s (1977) distinction between rote and smart instruments. Runeson introduced the notion of smart perceptual devices to describe perceptual mechanisms that directly register apparently complex or higher order informational variables, and he illustrated the notion with the analogy of a device called polar planimeter. In his model, perceiving is portrayed as the use of smart perceptual devices, such as the polar planimeter, by a central executive, who is to a large extent ignorant about the functioning of the devices. The strength of the notion of smart perceptual devices is that it explains perception without the need of a perceptual homunculus: The central executive is merely a cognitive (intentional) homunculus.

The direct learning theory can be said to concern smart learning devices. The difference between our view on smart learning devices and Runeson’s (1977) view on smart perceptual devices—apart from the fact that we do not have an equally illustrative analogy—is our emphasis on the self-adapting nature of the devices. If used to perceive,
or to detect higher order informational variables, the specificities of learning maintain the
learning device in shape for its purpose, by virtue of the design of the device and without
the user being aware of the details concerning the existence and functionality of these
specificities. Hence, as does the notion of smart perceptual device, the notion of smart
learning device needs only a cognitive homunculus: The learning processes themselves
are homunculus free.

Conclusions

Direct learning is a radical hypothesis because the entire explanatory burden is
carried by the specificity between information for learning and the to-be-explained
change with practice. Curiously, other learning theories need information concepts too.
Trial-and-error learning, for example, needs an information concept because trial-and-
error learners need to determine whether a particular change in performance is an
improvement or not. The difference is that, in trial-and-error learning, the information for
learning is detected after the change in performance and used, in an indirect manner, to
decide whether or not to undo the change. For direct learning, specificity between
information for learning and change in performance is all that is needed.

In the present article we have applied the theory to changes in variable use and
calibration. This does not mean that the theory is limited to this type of change in
performance. To the contrary, the theory is meant to apply to a wide variety of learning
phenomena. Because most learning theories need the concepts of change, information,
and specificity, and these are the only concepts needed in the direct learning theory, the
direct learning theory is very parsimonious. It therefore seems to be worthwhile to
explore which learning phenomena can be understood as direct learning. More boldly,
our claim is that, whatever the learning phenomenon under consideration, one should always carefully consider the direct learning hypothesis before one adopts less parsimonious hypotheses.
References


Author Note

David M. Jacobs, Facultad de Psicología, Universidad Autónoma de Madrid, and Center for the Ecological Study of Perception and Action, University of Connecticut; Paula L. Silva, Center for the Ecological Study of Perception and Action, University of Connecticut; Juan Calvo, Departamento de Matemática Aplicada, Universidad de Granada.

This material is based upon work supported by the National Science Foundation under Grant No. BCS-0339031 and by projects HUM2006-11603-C02-02 and MTM2005-02446 of the Spanish Ministry of Education and Science. We also thank Bill Mace, Claire Michaels, Jorge Ibañez, John van der Kamp, and two anonymous reviewers for helpful discussions and comments on earlier versions of the article.

Correspondence may be addressed to David M. Jacobs, Facultad de Psicología, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, España. Electronic mail may be sent to david.jacobs@uam.es.
Footnotes

1 We used an axis through the hand-held end of the object and perpendicular to the central rod of the objects for these definitions (look ahead to Figure 1 for an illustration of the objects).

2 In this equation, the values of $I_3$ and $\int \rho(s) \delta(s) \delta_x dV$ were both scaled to our response scale (see method section), for each value of $x$, before they were combined into a compound variable.

3 It might be interesting to note the following contrast between theories that are based on changes in variable use (e.g., the current application of direct learning) and theories that attribute changes in performance to changes in hypothetical algorithms that process (or combine) unchanging input variables (e.g., cue-combination, Bayesian inference). Whereas the former class of theories predict that perceptual learning goes together with changes in exploratory movements (arm movements, eye movement, head movements, etc.), the latter class of theories do not (see Arzamarski, Isenhower, Kay, Turvey, & Michaels, 2009, for more detail on this observation).

4 To appreciate this latter claim, imagine a point mass $M$, on an otherwise weightless rod, at distance $D$ from the axis of rotation. In such a situation, $SM = DxM$, which means that $SM/M = DxM/M = D$. In other words, the ratio of static moment and mass is related to the distance between the axis of rotation and the average location of the mass.
Additional analyses with regard to the slope of the function $f$ did not lead to noteworthy results. This might be related to the fixed response-scale that was used in the experiment.

It is possible to experimentally distinguish information-based or direct and gradient-based or indirect accounts. For instance, one can determine the locus of a participant in real time and add random (non-informative) error to the feedback depending on the current locus. This allows one to dissociate gradients and information for learning (Jacobs, Ibanez, & Travieso, 2009).

Two critical points are considered to be isomorphic if and only if an isomorphism can be defined between them. Our mathematical concept of isomorphism can intuitively be understood as the one of a sufficiently regular bijective function that maps the vectors of one vector field on the vectors of the other field. We will not aim to present a more precise interpretation of the term sufficiently regular that is consistent with the presented classification of critical points. Note, however, that classifications are determined by the definitions of the used isomorphisms and of the systems that one studies. For example, if one uses a more restrictive notion of isomorphism, one obtains a classification with more numerous classes of equivalence, and vice versa. Mathematical theorizing often consists of obtaining those definitions of systems under study and of the isomorphisms between them that lead to elegant classifications. The following sections are inspired by this type of reasoning.
Examples of differential manifolds include surfaces embedded in Euclidean spaces and Euclidean spaces themselves. Restricting the theory to such examples is sometimes convenient because it allows one to use concepts defined with the usual Euclidean metric (cf. Definition 21).

Other authors argue that individuals differ in how they learn, which seems to imply the rejection of Conjectures 2 and/or 3 (e.g., Withagen & Wermeskerken, 2009). We prefer not to consider individual differences at this level in order to aim for an understanding of learning that is as general and lawful as possible.

Many yet other classifications are possible. For example, Classification 2 is in fact a reinterpretation of Classification 1a, and one might likewise obtain a reinterpretation of Classification 1b. Second and more interestingly, direct learning systems might often be described with multiple specificities of action. Just to give an example, imagine that, in catching, the overall locomotion and the hand movement relative to the body are portrayed as being controlled independently of each other, on the basis on different informational variables. Each of these specificities of action might then be embedded in multiple specificities of learning. Taking this into account leads to classifications with more numerous classes of equivalence. As an aside, the latter portrayal raises the following question: Does the information for learning for one specificity of action depend on the current state of the other specificities of action? This question seems to be especially interesting for developmental issues, given that young individuals might typically function in ways that may be considered as far from optimal.
11 Note that local task ecologies in which different individuals perform particular actions (e.g., catch objects) differ in part because the individuals themselves select the task ecologies. For example, they perceive which objects are catchable and/or choose which objects to catch. This selection affects the usefulness of the variables implicated in the control of catching, the information for learning, and hence the attunement process. Likewise, the variables that are used for the control of the catching affect the task situation and the attunement with regard to the catchability judgments. Observe the similarities with the circular causality indicated in Footnote 10.

12 To give another ecologically-motivated example, one could consider spaces each point of which represents, say, a coordinative structure (Fowler & Turvey, 1978), and study whether learning goes together with movements through such spaces, whether or not such movements are specific to information for learning, etc.
Table 1

*Set of objects used in Experiment 1*

<table>
<thead>
<tr>
<th>Object #</th>
<th>Added Mass (g)</th>
<th>Position on Central Rod (cm)</th>
<th>Position on Crossbars (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80.4</td>
<td>37.5</td>
<td>21.0</td>
</tr>
<tr>
<td>2</td>
<td>87.6</td>
<td>37.5</td>
<td>9.0</td>
</tr>
<tr>
<td>3</td>
<td>73.2</td>
<td>37.5</td>
<td>21.0</td>
</tr>
<tr>
<td>4</td>
<td>66.0</td>
<td>37.5</td>
<td>9.0</td>
</tr>
<tr>
<td>5</td>
<td>51.6</td>
<td>37.5</td>
<td>21.0</td>
</tr>
<tr>
<td>6</td>
<td>44.4</td>
<td>37.5</td>
<td>9.0</td>
</tr>
<tr>
<td>7</td>
<td>30.0</td>
<td>37.5</td>
<td>21.0</td>
</tr>
<tr>
<td>8</td>
<td>37.2</td>
<td>37.5</td>
<td>9.0</td>
</tr>
<tr>
<td>9</td>
<td>87.6</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td>10</td>
<td>80.4</td>
<td>22.5</td>
<td>9.0</td>
</tr>
<tr>
<td>11</td>
<td>66.0</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td>12</td>
<td>73.2</td>
<td>22.5</td>
<td>9.0</td>
</tr>
<tr>
<td>13</td>
<td>44.4</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td>14</td>
<td>51.6</td>
<td>22.5</td>
<td>9.0</td>
</tr>
<tr>
<td>15</td>
<td>37.2</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td>16</td>
<td>30.0</td>
<td>22.5</td>
<td>9.0</td>
</tr>
<tr>
<td>17</td>
<td>58.8</td>
<td>30.0</td>
<td>15.0</td>
</tr>
<tr>
<td>18</td>
<td>58.8</td>
<td>30.0</td>
<td>15.0</td>
</tr>
</tbody>
</table>

*Note.* The added mass refers to the mass of each of the four attached metal weights.
Table 2

*Average correlations between judgments and mass \((r_{\text{mass}})\) and between judgments and static moment \((r_{\text{sm}})\) for the different groups of participants.*

<table>
<thead>
<tr>
<th></th>
<th>Mass Group ((n=8))</th>
<th>Static Moment Group ((n=9))</th>
<th>Control Group ((n=4))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest</strong></td>
<td>0.64</td>
<td>0.69</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>Posttest</strong></td>
<td>0.80</td>
<td>0.69</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Pretest</strong></td>
<td>0.83</td>
<td>0.80</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>Posttest</strong></td>
<td>0.78</td>
<td>0.88</td>
<td>0.91</td>
</tr>
</tbody>
</table>

\(r_{\text{mass}}\)

\(r_{\text{sm}}\)
Table 3

Sets of objects used in Experiment 2. M = mass of each of the four attached metal weights; R = position on central rod; and C = position on crossbars.

<table>
<thead>
<tr>
<th>Object #</th>
<th>Fast Set M (g)</th>
<th>R (cm)</th>
<th>C (cm)</th>
<th>Slow Set M (g)</th>
<th>R (cm)</th>
<th>C (cm)</th>
<th>Test-Phase Set M (g)</th>
<th>R (cm)</th>
<th>C (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>66.0</td>
<td>22.5</td>
<td>9.0</td>
<td>30.0</td>
<td>22.5</td>
<td>9.0</td>
<td>30.0</td>
<td>22.5</td>
<td>9.0</td>
</tr>
<tr>
<td>2</td>
<td>73.2</td>
<td>22.5</td>
<td>21.0</td>
<td>37.2</td>
<td>22.5</td>
<td>21.0</td>
<td>44.4</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td>3</td>
<td>80.4</td>
<td>22.5</td>
<td>9.0</td>
<td>44.4</td>
<td>22.5</td>
<td>9.0</td>
<td>66.0</td>
<td>22.5</td>
<td>9.0</td>
</tr>
<tr>
<td>4</td>
<td>87.6</td>
<td>22.5</td>
<td>21.0</td>
<td>51.6</td>
<td>22.5</td>
<td>21.0</td>
<td>80.4</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td>5</td>
<td>30.0</td>
<td>37.5</td>
<td>9.0</td>
<td>66.0</td>
<td>37.5</td>
<td>9.0</td>
<td>37.2</td>
<td>37.5</td>
<td>9.0</td>
</tr>
<tr>
<td>6</td>
<td>37.2</td>
<td>37.5</td>
<td>21.0</td>
<td>73.2</td>
<td>37.5</td>
<td>21.0</td>
<td>51.6</td>
<td>37.5</td>
<td>21.0</td>
</tr>
<tr>
<td>7</td>
<td>44.4</td>
<td>37.5</td>
<td>9.0</td>
<td>80.4</td>
<td>37.5</td>
<td>9.0</td>
<td>73.2</td>
<td>37.5</td>
<td>9.0</td>
</tr>
<tr>
<td>8</td>
<td>51.6</td>
<td>37.7</td>
<td>21.0</td>
<td>87.6</td>
<td>37.7</td>
<td>21.0</td>
<td>87.6</td>
<td>37.7</td>
<td>21.0</td>
</tr>
<tr>
<td>9</td>
<td>30.0</td>
<td>22.5</td>
<td>9.0</td>
<td>30.0</td>
<td>22.5</td>
<td>9.0</td>
<td>30.0</td>
<td>22.5</td>
<td>9.0</td>
</tr>
<tr>
<td>10</td>
<td>44.4</td>
<td>22.5</td>
<td>21.0</td>
<td>44.4</td>
<td>22.5</td>
<td>21.0</td>
<td>44.4</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td>11</td>
<td>66.0</td>
<td>22.5</td>
<td>9.0</td>
<td>66.0</td>
<td>22.5</td>
<td>9.0</td>
<td>66.0</td>
<td>22.5</td>
<td>9.0</td>
</tr>
<tr>
<td>12</td>
<td>80.4</td>
<td>22.5</td>
<td>21.0</td>
<td>80.4</td>
<td>22.5</td>
<td>21.0</td>
<td>80.4</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td>13</td>
<td>37.2</td>
<td>37.5</td>
<td>9.0</td>
<td>37.2</td>
<td>37.5</td>
<td>9.0</td>
<td>37.2</td>
<td>37.5</td>
<td>9.0</td>
</tr>
<tr>
<td>14</td>
<td>51.6</td>
<td>37.5</td>
<td>21.0</td>
<td>51.6</td>
<td>37.5</td>
<td>21.0</td>
<td>51.6</td>
<td>37.5</td>
<td>21.0</td>
</tr>
<tr>
<td>15</td>
<td>73.2</td>
<td>37.5</td>
<td>9.0</td>
<td>73.2</td>
<td>37.5</td>
<td>9.0</td>
<td>73.2</td>
<td>37.5</td>
<td>9.0</td>
</tr>
<tr>
<td>16</td>
<td>87.6</td>
<td>37.7</td>
<td>21.0</td>
<td>87.6</td>
<td>37.7</td>
<td>21.0</td>
<td>87.6</td>
<td>37.7</td>
<td>21.0</td>
</tr>
<tr>
<td>17</td>
<td>58.8</td>
<td>30.0</td>
<td>15.0</td>
<td>58.8</td>
<td>30.0</td>
<td>15.0</td>
<td>58.8</td>
<td>30.0</td>
<td>15.0</td>
</tr>
<tr>
<td>18</td>
<td>58.8</td>
<td>30.0</td>
<td>15.0</td>
<td>58.8</td>
<td>30.0</td>
<td>15.0</td>
<td>58.8</td>
<td>30.0</td>
<td>15.0</td>
</tr>
</tbody>
</table>
Figure Captions

*Figure 1.* Tensor object.

*Figure 2.* Distributions of the informational variables used by participants in the mass group (left panels), static moment group (middle panels), and control group (right panels), in each block of trials (Pretest, Practice 1, Practice 2, Practice 3, and Posttest). M = locus corresponding to mass. SM = locus corresponding to static moment.

*Figure 3.* Vector fields corresponding to Eqs. 3 and 4 for the mass group (upper panel) and static moment group (middle panel), along with the standard error ellipses for these groups in the pretest (dashed outline) and posttest (continuous outline). The bottom panel shows the standard error ellipses for the control group.

*Figure 4.* Vectors that indicate the change in variable use from block of trials to block of trials for participants in the mass group (upper panel), static moment group (middle panel), and control group (lower panel).

*Figure 5.* Predicted mass judgments as a function of candidate informational variables for Object 10. The dashed line-segments indicate the informational variables that go together with hypothetical judgments of 20 and -15.
Figure 6. x-Loci for the mass group (circles), static moment group (diamonds), and control group (squares), averaged per block of trials and over all participants.

Figure 7. Correlation between trial-to-trial change in x-locus (left-hand side of Eq. 7) and the informational candidate (right-hand side of Eq. 7) for different numbers of used trials (m in Eq. 6) for the mass group (left panel, continuous line) and the static moment group (right panel, continuous line), and for the control group with error defined with regard to mass (left panel, dashed line) and with error defined with regard to static moment (right panel, dashed line).

Figure 8. Correlation between trial-to-trial change in calibration (left-hand side of Eq. 9) and the informational candidate (right-hand side of Eq. 9) for different numbers of used trials (m) for the mass group (left panel, continuous line) and the static moment group (right panel, continuous line), and for the control group with error defined with regard to mass (left panel, dashed line) and with error defined with regard to static moment (right panel, dashed line).

Figure 9. Correlation between the estimated calibration constant on trial n+1 and the error observed on trials n, n-1, n-2, and n-3, for the mass group (left panel, continuous line) and the static moment group (right panel, continuous line), and for the control group with error defined with regard to mass (left panel, dashed line) and with error defined with regard to static moment (right panel, dashed line).
Figure 10. Vector fields corresponding to Eqs. 3 and 4 for the fast group (upper panel) and slow group (lower panel) along with the standard error ellipses for the two groups in the pretest (dashed outline) and posttest (continuous outline). M = locus corresponding to mass. FB = locus corresponding to feedback variable. SM = locus corresponding to static moment.

Figure 11. Correlation surfaces for the fast group (upper graph) and slow group (middle graph) and the pretest and posttest loci of the fast group presented as probability density functions (lower graph).

Figure 12. Examples of vector fields with flow curves.
Figure 1
Figure 2

The figure shows the x-locus and y-locus for Mass, Static Moment, and Control conditions across Pretest, Practice 1, Practice 2, Practice 3, and Posttest. The loci are displayed in different quadrants, indicating changes in position and movement patterns over time.
Figure 3

Mass Group

Static Moment Group

Control Group
Figure 4

- Mass Group
- Static Moment Group
- Control Group

(x-Locus vs y-Locus plot for different groups)
Figure 5
Figure 6
Figure 7

The image shows two graphs: one for Mass and one for Static Moment. Each graph plots the correlation against the number of trials (m). The graphs compare Mass Group vs. Control Group and Static Moment Group vs. Control Group. The correlation values range from 0.1 to 0.5.
Figure 8

![Graph showing the correlation between mass and static moment over the number of trials. The graph compares Mass Group and Control Group, as well as Static Moment Group and Control Group. The correlation decreases as the number of trials increases.]
Figure 9

The figure presents a comparison of correlation between Mass and Static Moment across different trial groups. The left panel shows the Mass group with a solid line and the Control group with a dashed line. The right panel depicts the Static Moment group with a solid line and the Control group with a dashed line. The x-axis represents the trials labeled as n, n-1, n-2, and n-3, while the y-axis represents the correlation values ranging from -0.15 to 0.10.
Figure 10

Fast Group

Slow Group

Pretest
Posttest

M FB SM

x-Locus

y-Locus
Figure 11

Fast Group: Correlation Surface

Slow Group: Correlation Surface

Fast Group: Probability Density Functions
Figure 12

Stable Node

Stable Spiral