The Assessment and Modeling of Perceptual Control
A Transformation in Research Methodology to Address the Replication Crisis

Warren Mansell
School of Health Sciences, University of Manchester, UK

Vyv Huddy
Research Department of Clinical, Educational and Health Psychology, University College London, University of London, UK

RUNNING HEAD: PERCEPTUAL CONTROL MODELING

Address for correspondence:
Dr Warren Mansell
Reader in Clinical Psychology
CeNTrUM (Centre for New Treatments and Understanding in Mental Health), Division of Psychology and Mental Health, School of Health Sciences, Faculty of Biology Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, 2nd Floor Zochonis Building, Brunswick Street, Manchester, M13 9PL, UK
Email: warren.mansell@manchester.ac.uk
Tel: +44 (0) 161 275 8589
Website: http://www.psych-sci.manchester.ac.uk/staff/131406

10.1037/gpr0000147

Abstract

Replication in the behavioral sciences is a matter of considerable debate. We describe a series of fundamental interrelated conceptual and methodological issues with current research that undermine replication and we explain how they could be addressed. Conceptually, we need a shift (1) from verbally described theories to mathematically specified theories, (2) from lineal stimulus-cognition-response theories to closed-loop theories that model behavior as feeding back to sensory input via the environment, and (3) from theories that 'chunk' responses to theories that acknowledge the continuous, dynamic nature of behavior. A closely related shift in methodology would involve studies that attempt to model each individual's performance as a continuous and dynamic activity within a closed-loop process. We explain how this shift can be made within a single framework – perceptual control theory - that regards behavior as the control of perceptual input. We report evidence of multiple replication using this approach within visual tracking, and go on to demonstrate in practical research terms how the same overarching principle can guide research across diverse domains of psychology and the behavioral sciences, promoting their coherent integration. We describe ways to address current challenges to this approach and provide recommendations for how researchers can manage the transition.
Keywords: experimental design; replicability; computational models; closed-loop; negative feedback control; perceptual control theory

Conflicts of Interest

The authors declare no conflicts of interest with respect to the authorship or the publication of this article.

Contributions of Authors

Both authors contributed equally but the first author took the lead in writing the article

The Assessment and Modeling of Perceptual Control
A Transformation in Research Methodology to Address the Replication Crisis

The replication crisis in psychology is in little doubt (Pashler & Wagenmakers, 2012). There is a similar unease in the life sciences more widely that has existed for some time (Ioannidis, 2005) and has not been resolved (Horton, 2015). Large-scale replication efforts have had disappointing results. The most widely publicized has been the Open Science Collaboration (OSC) (2015) that found an overall replication rate at 36% with many effects much smaller than the original reports. There has been a vast amount of commentary on this contentious topic and the debate has generated a range of solutions. These have often highlighted the practices of research in the behavioral sciences with the emphasis on transparency and integrity (Ioannidis, 2005; Nosek, 2012; Wagenmakers, 2012; Stevens, 2017). In this view, the sole change would be that the traditional methodological paradigm would be executed more rigorously.

Previous commentaries on replication have clarified the commonly discussed statistical issues with sampling error, multiple testing, and the issues with replication of small effect sizes (e.g. Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson, & Munafò, 2013). Yet, rarely have commentators taken issue with the conceptual basis for research designs, and the *fundamental* statistical and methodological assumptions that are made. In this article, we explain how a series of interweaving conceptual and methodological issues will continue to undermine the replication of psychology experiments unless they are addressed. We then introduce a new research paradigm – based on the ‘control of perception’ - that has potential to address all of these issues, and we provide examples of studies within human performance, animal behavior, clinical, social, organizational and developmental psychology. We describe the challenges ahead of transitioning to a wholly different perspective on psychology, and a roadmap of how it may be achieved.

Conceptual problems with prevailing assumptions

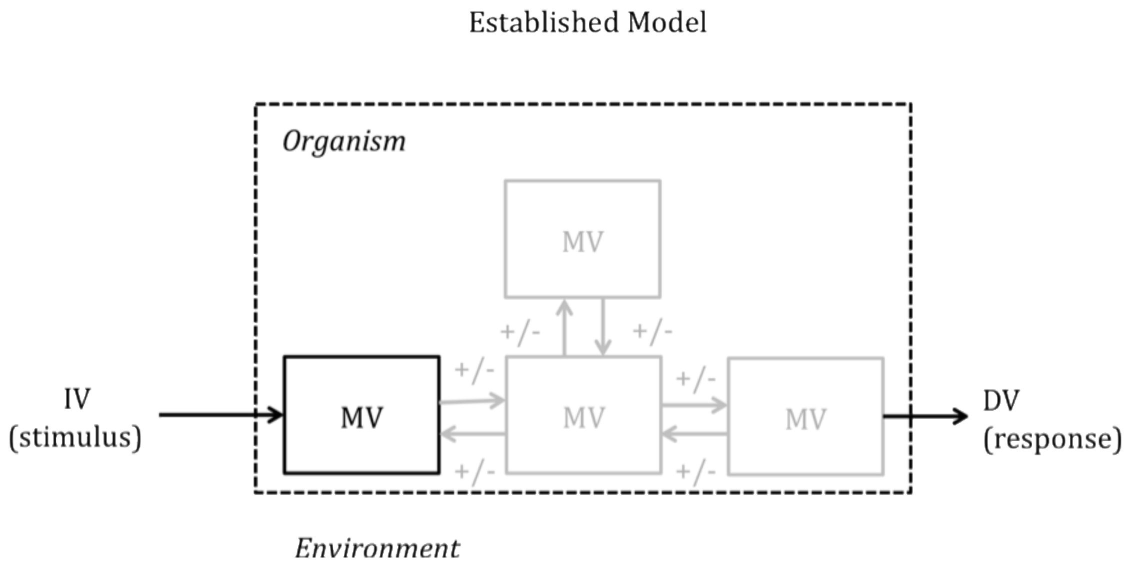
A1. Researchers ordinarily formulate their theories verbally and not mathematically.

The vast majority of behavioral research is interpreted using theories of a phenomenon that are communicated verbally. As Rodgers (2010) points out “our language is... a model” (p. 1); these verbal theories are the models often spoken about in psychological science, such as the generate-recognize model of memory (Anderson & Bower, 1973) and the planning control model (Glover, 2004). These examples were chosen as they are both known by the term 'model' and yet are specified only verbally so the description as a *model* can only be metaphorical. Language is, however, inherently ambiguous, meaning there is a constant risk of disagreement of the implications of a verbally stated theory. Arguably, even the most reliable findings can be interpreted in different ways, entailing uncertainty when attempting to confirm the replication of previous findings.

A2. *Theories are typically oversimplified by specifying that variations in the independent variable (IV) cause variations in the dependent variable (DV).*

Figure 1 shows typical IV-DV approach. It is assumed that manipulating the IV changes some aspect of the stimuli used in the experiment and the DV is the variation in the measured response. Yet, there is a recognition that individuals act as agents that act dynamically within their environment such that the causal pathway is not a simple one-way process from stimulus to response (El Hady, 2016; Schlesinger & Parisi, 2001; Smith & Conrey, 2007). A participant’s behavior, alongside unmeasured disturbances in the environment, has a feedback effect on their sensory input. This feedback effect was noted as far back as the nineteenth century by John Dewey (1896), “the motor response determines the stimulus, just as truly as sensory stimulus determines movement” (p. 363). One important example is during eye movements, which entail that the ‘stimuli’ perceived are differently from moment to moment (Land & Furneaux, 1997). Given that many psychological theories do not incorporate sensory feedback and unseen disturbances within the model, their findings are unlikely to be replicable.

Figure 1. A generic diagram of an established cognitive or behavioral model. The diagram shows the potential for an array of sequential and top-down inhibitory and excitatory processes in gray. The feedback effect of the response on the sensory effect of the stimulus is rarely shown in diagrams and if it is it is treated as though it was part of a sequential process. Yet the feedback effect is always present, though not necessarily on the sensory effect of the stimulus (IV) used in the experiment. IV = independent variable; DV = dependent variable; MV = mediating variable (mental process)



A3. Attempts to isolate discrete behaviors are often arbitrary.

The IV-DV model attempts to link discrete stimuli with discrete responses, or sequences of discrete stimuli and responses. However, behaviors are not discrete in themselves. They are often embedded in other, ongoing processes. Consider the ‘behavior’ of opening a car door. It could be defined, and therefore measured, as: the experimenter’s measurement of the door being opened; the movement of the door towards an opened state; the arm movements necessary to open the door; the muscular forces necessary to open the door; the motor signals sent to the muscles that move the arms to open the door. Importantly, none of these definitions are ‘wrong’, but the fact that there are at least five different plausible definitions shows how arbitrary any one of these definitions can be. With very little consensus in this matter of what is defined as behavior, there is wide scope for differences in interpreting what counts as the replication of behavior.

Statistical and methodological problems with prevailing assumptions

B1. Variation between individuals means that group averages do not apply to any one individual.

Most psychology studies collect data from groups of participants and they report summary statistics of average performance. This approach is limited because the ultimate purpose of a psychological theory is to describe the workings of an individual, and not of a group. One notable example of where group data has led to an misleading conclusion is the large body of research leading to the conclusion that there is a ‘learning curve’ (Gallistel, Faurhurst, & Balsam, 2004); analysis of individual animals reveals that rather than a curve function, performance improves from pre-training as discrete step-like increases in performance.

Indeed, group statistics can lead to erroneous conclusions about relationships between variables that are directly opposite to the known relationship within computational models of individuals (Powers, 1990). Powers (1990) constructed individual computational agents whose level of effort was increased when reward *decreased*. Each of these agents had a parameter of reward sensitivity that was set by random for each individual. When plotting the level of reward by the level of effort for each individual in a large sample, there was a significant positive correlation between increasing reward and increasing effort. Thus, the reverse relationship was observed within the group to that which had been implemented within the individual.

B2. The way that individual variation in behavior is analyzed adds to uncertainty.

It is typically noted that measured behavior is variable in experimental tasks (Bell, 2014), and this is especially true in single-case designs where data are not averaged across participants or across repeated measures (Normand, 2016). Behavior may also vary on a trial-by-trial basis across seemingly identical situations in an experiment (Gluth & Rieskamp, 2017). Many models of behavior do not account for any form of variability between individuals and ignore these trial-by-trial fluctuations. One method is to seek to model average performance across trials to obscure this difficulty. Indeed, some regard variable behavior as evidence of intrinsic random noise; this was the policy of the early behaviorists who "solved the problem by attributing the unpredictability of behavior to a universal property of living organisms: variability" (Powers, 1973, p5). Yet, to the degree that individual variation in behavior is *not* random but is due to an as-yet-unspecified mechanism, replication will be unnecessarily compromised.

B3. "Open loop" research designs and laboratory settings do not represent 'normal' behavior outside the laboratory

We made the case earlier that organisms are 'closed loop'. Whatever the design of an experiment, variations in the IV cannot be the proximal cause of organism's actions because these events occur at a *distal* location in the organism's environment (e.g. the appearance of flashing lights, pictures or sounds). These events produced by the IV have *proximal* effects on the participant via the excitation of sensory nerves (Marken, 1997; Powers, 1978). Moreover, the behavior of the organism – measured by the DV – *also has sensory effects*. For example, in any experiment where a stimulus offset is triggered by the response, the duration of stimulus presentation – a proximal sensory effect - is influenced by reaction time (the DV in many cases). This means that the proximal sensory effects of the experimental circumstance are a combination of both the experimental manipulation (IV) and participant's behavior (DV) (Marken, 1997, 2013). This process can be ongoing, and simultaneous and does not necessarily proceed in a sequence of actions and events (Powers, 1992). Sensory effects have often been reframed as consequences or reinforcements (Baum, Reese, & Powers, 1973). However, sensory effects are the *combined* effect of behavior *and* environmental disturbances.

Researchers often assume that an open-loop design is necessary to study behavior accurately, even though they acknowledge that in normal circumstances humans, and animals' are closed-loop in nature (e.g. Heisenberg & Wolf, 1992). This assumption is implicit in the design of the classic reaction-time task that presents stimuli as a distinct event and measures a response. Some human studies claim to be open loop when sensory input from one channel is obstructed (e.g. reaching in the dark; Henriques et al., 1998). In animal studies, the design is often more elaborate and involves using an apparatus to immobilize the animal to convert the design to 'open loop' (Heisenberg & Wolf, 1992). We propose that if organisms are operating as closed loop systems in most cases, attempts to generate an open-loop design are at the least artificial, and at the worst, misleading because humans and other organisms are likely to find ways to circumvent the procedure (e.g. by using an alternative sensory modality). These adaptive reactions are likely to be inconsistent and introduce variability into an experimental procedure that reduces the capacity for replication.

Research on body movements in the context of affect provides one extended example within the field of social and clinical psychology of how open loop studies have led to non-replicated and mixed findings. Based on embodiment theory, it has been proposed that there are inherent bodily movements associated with certain affective stimuli (Laham, Kashima, Dix, & Wheeler, 2015). A series of open loop studies have tested whether positive as opposed to negative affective stimuli are associated with the response of biceps flexion rather than extension, because biceps extension is conceived as biological tendency to push away aversive stimuli. A meta-analysis of 68 independent effect sizes revealed a significant but weak effect (Laham, Kashima, Dix, & Wheeler, 2015). Further analyses revealed that the effect is actually reversed by framing biceps extension as approach and flexion as avoidance, rather than framing them as pulling and pushing a stimulus in relation to the self. The authors of the meta-analysis concluded that participants attempt to *keep negative stimuli at a further distance from oneself* than positive stimuli, regardless of the exact muscle movements involved. Thus, the attempts to replicate a specific stimulus-response mapping have failed, in place of evidence that closed loop control of perceived distance may be the consistent feature shared across studies.

B4. Studies of mediators and "mechanisms of change" tend to be subject to the above issues

At times, groups of researchers can conflict for many years over what is the 'correct' theory of a psychological phenomenon. This can obscure the possibility that different theories may apply to different individuals within any sample. For example, this is evident when participants spontaneously employ different strategies in a navigation task (Iaria et al. 2003). The separation of these groups indicates that neither strategy mediates the relationship between task instructions and performance across all participants. Indeed, participants were also shown to change strategies during the task, meaning neither strategy accounted for the behavior of any individual participant. A special version of this issue is Simpson's Paradox, where combining different groups of participants may show the reverse effect of the two groups studied independently (Blyth, 1972). Often, a highly integrative research design and the consideration of multiple moderators of an experimental effect are used to attempt to discover such relationships within the data (e.g. Colquitt, Scott, Judge, & Shaw, 2006).

A prominent example of where group comparisons can lead to erroneous conclusions is within the randomized controlled trials used to compare different forms of psychological therapy. Whilst these trials can demonstrate the relative superiority of a certain intervention, they cannot, on their own, provide any test of the theory informing the therapy. Studies of the 'mechanism of change' of psychological therapies may use statistical analyses to examine mediating variables (MVs; e.g. Warmerdam, van Straten, Jongsma, Twisk, & Cuijpers, 2010). Yet, these patterns of relationships across individuals are prone to the same errors as illustrated above. Indeed, there is a wide individual variation in the outcomes and the temporal profile of psychological change that are rarely assessed (Hayes, Laurenceau, Feldman, Strauss, & Cardaciotto, 2007). In short, using group statistics to infer a mechanism of change is prone to errors that reduce the likelihood of replication.

The intertwined nature of conceptual, statistical and methodological problems

The nature of the problems describe above are reciprocally related. Most researchers recruit groups of participants to carry out open-loop experiments and analyze the effects of discrete stimuli on discrete responses. This approach to research inevitably constrains these researchers to only test simple IV-DV hypotheses in spite of the conceptual shortcomings of the theories of this kind. Similarly, if theories are limited to those that specify only direct pathways and fail to consider closed-loop feedback, they will be constrained to the traditional statistical designs with the errors and uncertainties we have described. Some IV-DV protocols demonstrate high levels of replication across groups of participants but even in these cases it is rare for all individuals in a sample. We will demonstrate below that closed-loop methods hold potential to increase the bar to replicating in every case, and not only in every study.

An alternative approach to conceptualization and methodology

Following from the above analysis, a future of replicable research requires that each of the above conceptual and methodological problems is addressed. It is unlikely to be sufficient to simply address some of these issues because any one of them can undermine replicability. Specifically, it will require all of the following within a new approach:

A1. A mathematically specifiable psychological theory

When a theory can be specified mathematically, it removes the uncertainty surrounding verbal terms and their various interpretations (McClelland, 2014). It also allows the nature of the relationships between the elements of a theory to be specified. This in turn allows a computational model to be constructed and the pattern of expected data can be specified and tested directly against the real world data. This greatly reduced uncertainty enhances the capacity for replication. Arguably one of the most successful mathematically specified theories is evolution by natural selection (Mansell, Carey, & Tai, 2015). Within psychology, as we have argued, they are rarer. Those that do exist are most easily found in cognitive science, such as the General Context Model (Nosofsky, 1986) - a theory of object classification. Broader mathematical theories in psychology and neuroscience are more limited but one contemporary example is the free energy principle (Friston, 2010)[1].

A2. A theory that acknowledges and incorporates the closed-loop nature of humans and other organisms

Closed-loop theories treat humans and other animals as they are in their natural environment, in which any action leads to a change in sensory input to the organism. There is no artificial point where an organism stops acting to control its perceptions. The experimental situation operations to determine what perceptual variables might be under control by the organism. Thus, there is no attempt to *control behavior* through a restraining apparatus. Instead, disturbances in the environment are introduced that the organism must counteract to meet its goals. The chances of replication are therefore enhanced because it allows participants greater freedom to adapt behavior to meet the demands of the experimental protocol rather than fixing their behavior into an arbitrary pattern.

A3. A theory that acknowledges the complex, continuous and dynamic nature of behavior

A theory that conceives of behavior as dynamic and continuous can inform a model to simulate the real system to generate data that can be assessed for its fit with data in the real world *as it is recorded*. The potential for replication is enhanced because information is not lost through imposing segmentation or smoothing of the data to remove variability. The variability in behavior is used to increase the chance that the perceptual variables can be identified and replicated in future.

B1: A methodology that attempts to model each individual's actions or performance

A single-case design, in which data from an individual is collected as repeated measures and replicated within the individual, can address many of the issues regarding inferential group statistics (Kazdin, 2011; Normand, 2016). Following from above, when a theory can be described mathematically, it can be reproduced as a computational model. The model yields a simulation of the behavioral phenomenon, in essence reproducing it, so that detailed observations and measurements may be made. A computational model (sometimes known as a *functional* model; Runkel, 2007) formalizes a hypothesis regarding the processes that occur within the individual within computational terms such that the behavior of the real system can be *simulated*. There are a number of standardized methods for comparing the suitability of competing models and the living system they attempt to emulate (Pitt, Myung, & Zhang, 2002). The most obvious metric is whether or not the model fits the observed data. It has been proposed that computational model testing of individual cases provides the most direct test of the theory underlying a therapy (Carey, Tai, Mansell, Huddy, Griffiths, & Marken, 2017) or underlying an organizational activity such as work scheduling (Vancouver, Weinhardt, & Schmidt, 2010). Yet other commentators rarely point out that matching a computer model of an individual with data collected *within that individual* is the ultimate test of a computational model (e.g. Smith & Conrey, 2007). It is critical to note that even models of the behavior of individuals within social groups require modeling of distinct individuals with distinct goals and preferences within that population, not of the 'group average' (McClelland, 2014; McPhail, Powers, & Tucker, 1992; Shoda et al., 2002).

Within the field of psychophysics, the behavior of individuals is well studied. For example, Steven's Power Law (Stevens, 1957) describes the relationship between the objective magnitude of a feature of the environment (e.g. the recorded decibels of a tone) and its subjectively perceived intensity (e.g. loudness). The exponent of power function that relates these two variables is found to be distinctive to individual participants and consistent over time (Logue, 1976). While there have been failures to establish the consistency of these individual-specific parameters over time (Teghtsoonian & Teghtsoonian, 1983), a range of studies have shown test-retest reliability in estimated parameters across a range of domains of perception beyond loudness, including perceptions as diverse as estimated distance (da Silva & Fukusima, 1986) and elastic stiffness (Nicholson, Adams, & Maher, 2000). The estimation of stable individual-specific parameters therefore seems achievable and can enhance replication, relative to attempting to replicate group-averaged data drawn from an attempt to randomly sample a population.

B2: A methodology that assesses and analyzes continuous variations in behavior

A robust approach is to systematically model variability in behavior before one concludes that the variation in behavior is inexplicable (Bell, 2014). It is important to note that we are not claiming that all variables in a psychology experiment *must* be continuous – there are clearly elements that are discontinuous, such as the experience of discrete *events*, the recognition of specific *configurations* of sensations as patterns, and the constructions of *categories* of features of the environment (Powers, 1973). Rather, we are proposing that researchers need to focus on measuring and modeling continuous variables because using this raw data *where it clearly occurs in the real world environment* will entail more information for enhancing replication, relative to artificially parsing continuous data into averaged, smoothed or categorized, chunks.

B3: A closed-loop methodology

Closed-loop task design allows individuals substantial control over their environment, and their sensory input, as in the real world. For example, this may involve interactions with quantitative coding of videos of natural environments, eye tracking across visual scenes and movement within a virtual reality environment. Yet, even traditional experimental tasks can be reconceptualized as closed-loop tasks; because organisms are closed loop by nature, this is the only way that they can engage with the task in practice. For example, Marken (2013) demonstrated how the traditional use of a button press to denote detection latency can be reconceptualized as a closed-loop design and shown to have a comparable level of fit to the open-loop design.

B4: A methodology that assesses how a mechanism of change operates within the individual

Many studies purport to measure the mechanism of change when in fact they measure the reported experiences of groups of people who have already undergone change (Carey et al., 2017). Rather, a methodology assessing change within the individual has the capacity for greater accuracy and therefore replicability, especially if compared against a computational model that incorporates an algorithm for the change process itself (Marken & Carey, 2015).

The Alternative Paradigm: Perceptual Control

Whilst there are approaches to psychological research that have provided one or more of the above solutions, we propose that *all* of these solutions need to be implemented to improve replication rates in psychological science. This is our reason for introducing perceptual control theory (Powers, Clark & McFarland, 1960a, b; Powers, 1973, 2005) and its associated methodological innovations. This theory, and its accompanying methodological framework, addresses all the difficulties we have outlined above in the following ways:

A1. PCT is a mathematically specifiable psychological theory

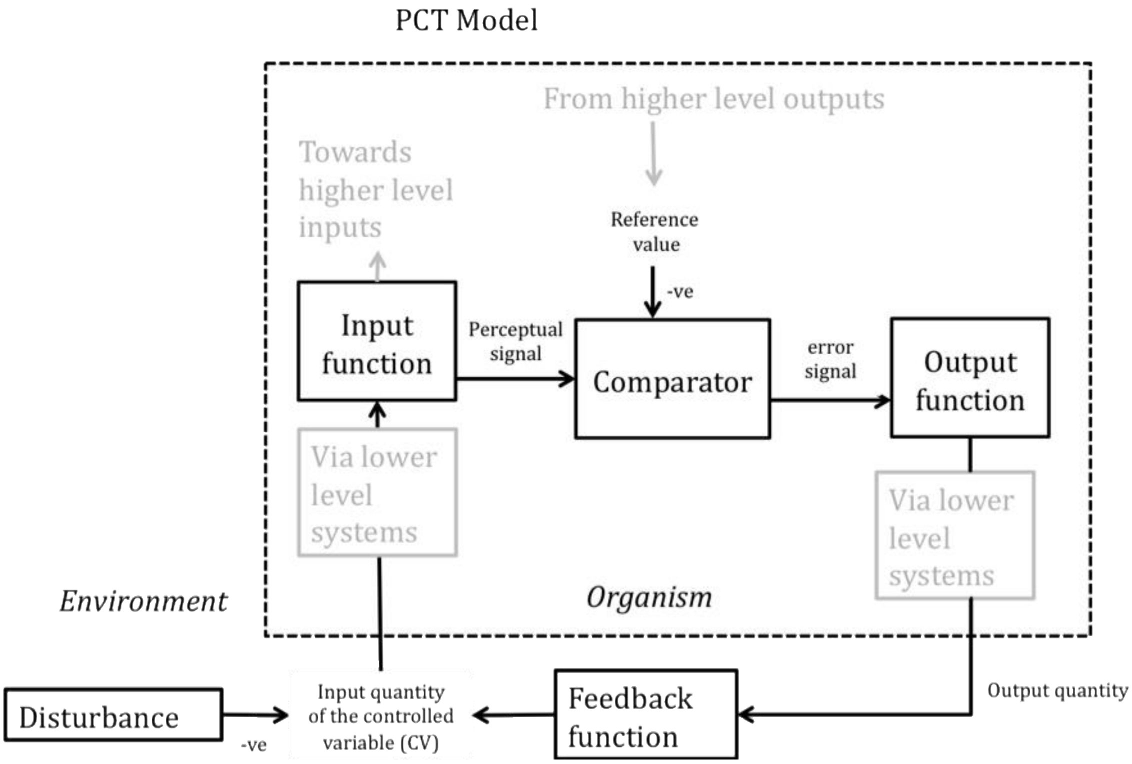
In a series of papers in the late 1970s Powers (1978 a-d) laid out in detail and defined the variables for functional models of a range of tasks. This included all the functions that transformed specific signals, as well as being precise about the location of input and output boundaries between the individual and the environment. Together, this entails a level of precision of model description that is necessary for effective replication. We will later demonstrate the level of replication that has been possible using these methods.

A2. PCT is a theory that acknowledges and incorporates the closed-loop nature of humans and other organisms

PCT provides an explanation of the phenomenon of control as a unique configuration of closed-loop negative feedback process (Powers, 1973). It regards behavior as a process of control that involves bringing perceived aspects of the environment to pre-specified (intended/purposive) states. Control is defined as the "achievement and maintenance of a preselected perceptual state in the controlling system, through actions on the environment that also cancel out disturbances" (Powers, 1973, 2005; p296).

The fundamental building block of this closed-loop process is shown in Figure 2. The *input function* extracts the aspect of the environment that is to be controlled. For example, in the tracking task explained later, the input function converts the sensory information from retinal cells to a perception of the position of the dot. The input function sends a *perceptual signal* that quantifies the currently perceived state of that aspect of the environment. In turn, it compares this value with *a reference value* (internal standard or goal; the source of this signal is explained below) for that variable – which could be 'on target' in this example. The discrepancy (or *error*) between the value inputted and the reference value drives *output signals*. These output signals are converted to actions (e.g. handle position) in the environment through the *output function*. For example, the error signal is typically amplified by a *gain* factor so that actions to reduce error are much greater than the error signal itself. In this way, even tiny adjustments can be made to keep on target with necessary force, such as when keeping a car in its lane during a storm.

Figure 2. The PCT control unit. Definitions of terms are provided in the text. The boxes denote functional operations that are applied to quantities within the environment or to signals within the organism. The dotted box denotes the organism-environment boundary. The minus sign denotes where a quantity is subtracted from the quantity passing around the loop. This single control loop is a functional simplification of a hierarchy of control loops that are represented in grey.



A3. PCT is a theory that acknowledges the complex, continuous and dynamic nature of behavior

There is no point within the PCT architecture at which 'behavior' can be defined as a specific output from the organism. Instead, it is vital to trace the pathway from neural signals to muscles, through the body and environment, considering the effects of other features of the environment, and back to the aspect of the environment (the *input quantity*) that is being controlled. Behavior is a property of the whole loop: organism and environment. The aspects of the environment that enable an individual to control the input quantity are termed the *feedback function*, whereas external influences on the input quantity are termed *disturbances*. They both combine to affect the input quantity directly. It is important to note that the individual does not perceive either the feedback function or the disturbances in order to control the input quantity. Only the perceptual signal is perceived. The environmental components are only ever noticeable via their effects on the controlled variable. Thus, there are important aspects of the environment that are not 'stimuli', in that the individual does not sense them directly. The individual merely needs to vary its outputs in a way that keeps the input quantity matching the reference for the state of that perception. As can be seen from the grey features within Figure 2, any control unit is actually one of many, and these are connected in a hierarchical network to control a variety of perceptual variables at any one time (Powers, 1973). For the sake of simplicity of explanation, we continue with the example of a single unit and elaborate on some of the more complex details later.

B1-4: PCT utilizes a methodology that attempts to model the individual's dynamic behavior as part of a closed-loop process

In PCT, the variable that the participant strives to keep at the reference value is termed the *controlled variable* (Marken, 1997). So, for example, during a reaction time experiment a participant may be attempting *to keep their environment free of aversive images* (e.g. Tolin, Lohr, Lee, & Sawchuk, 1999). The controlled variable would be the duration the image is displayed for and the reference value might be zero milliseconds for the duration of aversive images. Whatever the experimental manipulation, the presentation of aversive stimuli will influence any dependent variable that involves the display duration of the images; this allows an IV-DV relationship to be observed. However, it will only have an effect if the participant is controlling for a specific state of the sensory variable relevant to aversive images. Furthermore, as the relevance of the IV manipulations increase in relation to this controlled variable, then the effect size of the IV-DV relationship also increases. In the example, this could occur by increasing the vividness or size of the aversive images; larger or more vivid images provide a greater disturbance.

The formal method for identifying the controlled variable in an individual is called the Test for the Controlled Variable (TCV; Powers, 1973). The experimenter attempts to identify which perceptual variables the participant is controlling, and then to test this with precision, the experimenter may build and test a model that attempts to simulate the individual interacting with their environment. Only an extremely tight match of behavior between the model and the participant is counted as support for the model.

The TCV involves the following steps:

1. Hypothesize what variable is being controlled.
2. Apply disturbances that should affect the variable if the person was not controlling it.
3. Observe the effect of the disturbance. If the disturbance is resisted then the variable could be the CV. Try another disturbance to test this. Otherwise return to (1).
4. Once the experimenter can reliably create disturbances that are resisted then the CV is discovered.

The TCV can be applied across any domain of interest. The study is set up differently from a traditional IV-DV experimental design. First, the experimenter does not attempt to control all variables in the study; rather the experimenter aims to set up a design in which the *perceptual* variables that the participant controls can be *inferred* by the effect of their actions the controlled variable which is observed by the experimenter. Second, the experimenter does not present discrete ‘stimuli’ and measure discrete ‘responses’ to this stimulus. For example, in animal studies, the TCV involves recording the movements of free-moving animals rather than restricting movements and presenting stimuli to observe specific responses (Bell & Pellis, 2011; Barter et al., 2015). Instead, the experimenter needs to (a) characterize the continuous array of perceptual information that is available to the participant through their senses (b) identify or provide the means through which the participant has to control their perceptual input within the environment, and (c) identify or manipulate the disturbances to this control. Third, the TCV is aimed at modeling the *individual* rather than averaging across a group of individuals.

We will now illustrate how PCT is used to derive computational models of the individual and test them and we report evidence of the high levels of replicability of the fit of these models to human data. We begin with visual tracking in pure and applied settings, then introduce proximity control across a range of subdisciplines of psychology, then illustrate how higher order goals, and ultimately the more complex processes of learning, decision-making and reasoning may also be studied in closed-loop designs and modeled at the individual level.

Evidence for replication and robustness of PCT-informed methodology

1. Visual tracking

Powers (1973) describes the paradigmatic case of a closed-loop experiment as "the *tracking* experiment where a participant manipulates a control lever to cause a cursor - say a moving spot of light - to track a moving target (e.g. Taylor & Birmingham, 1948)" (p. 44). According to Powers (1973), "this is clearly a control task. [The participant] is trying to keep the spot and the moving target in a particular relationship, namely, on the target". If the dot moves from the target then the participant's response is always to reduce the difference via action. A range of visual tracking tasks have been tested, each requiring the control of a simple perceptual variable. They have involved control in one and two-dimensions, and movements determined by the user moving a computer mouse or operating a handle.

This task permits extremely reliable experimental effects in the form of the moment-by-moment intra-individual correlations obtained between the controlled variable (e.g. location of the cursor relative to a target), the output (e.g. the movements of the computer mouse), and the disturbance (e.g. the disturbance to the location of the target). In particular, it is found that the disturbance and the behavioral output of the individual have a very high negative correlation due to the fact that the individual’s action needs to counteract the disturbance in order to maintain control. Unlike typical experiments involving the behavior of humans or animals, the IV-DV relationship in tracking is extraordinarily high - in excess of .98 (Bourbon, 1996; Bourbon et al., 1990). In contrast, the controlled variable, even though it represents the input to the human participant, has a very low correlation with output or actions, thereby contradicting a lineal input-process-output model.

The next stage involves model generation, which reproduces the phenomenon of control manifest in any given task. In the case of simple tracking this is demonstrated by the model simulating the maintenance of the cursor on the target by appropriately countering disturbances. There is a training phase during which a computer model of one or more control units (see Figure 2) is constructed by estimating its parameters. For example, this involves inferences for the controlled variable (e.g. location of the cursor), its reference value (e.g. aligned exactly with the target) and the output function (e.g. the gain of the error signal). Then a test phase is carried out in which both the human participant and the computer model are presented with a disturbance pattern that has not been presented in previous trials. In each case, the agent needs to move the computer mouse by an amount that exactly tracks the target on the screen. In order to evaluate the accuracy of the computer model, a correlation coefficient is calculated for each participant by recording the position of the human's cursor and the position of the computer's cursor at each time point during the task, which typically around 16ms. A correlation of 1.0 would indicate perfect correspondence between the positions of both the human and computer. An example of the task itself is displayed in Figure 3.

Figure 3. A) Experimental set up with computer model and screen: r = reference signal, p – perceptual signal, C = cursor position, T = target position, e = error, o = output signal; B) Experimental set up from the view of the participant, C) Results typical of human tracking over one experimental run. Reproduced from Parker et al. (2017), pending permission from Sage publications.

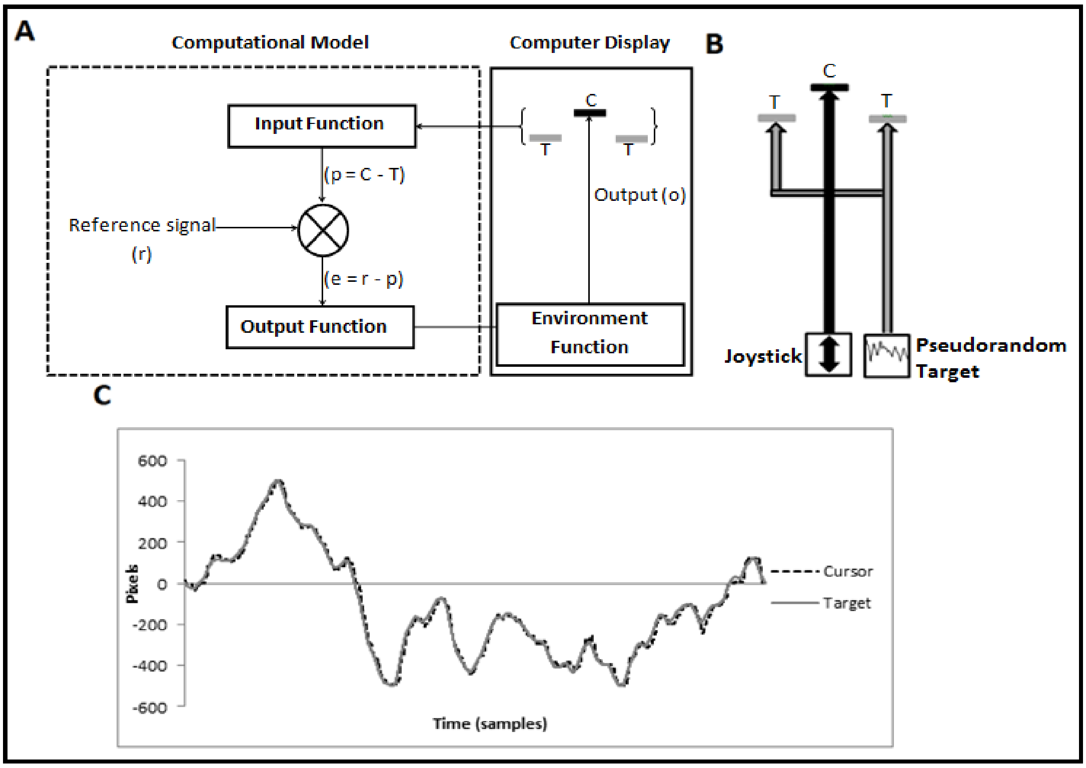


Table 1. Published replications of high correlations between the cursor movements generated by a PCT computer models and the recorded cursor movements generated by the individual

Task and Study	*Pearson correlation*
Relative location in one dimension (Pursuit)	
Bourbon & Powers (1999)	
Experiment 1	r = .989
Experiment 2	r = .981
Experiment 3	.969 ≤ r ≤ .996
Marken (2013)	r = .995
Marken and Powers (1989)	
Experiment 1	r = .987
Experiment 2A	r = .95
Experiment 2B	r = .95
Experiment 3	r = .94
Experiment 4	r = .97
Bourbon (1989)	
Experiment 1	.991 ≤ r ≤ .996
Experiment 2	.991 ≤ r ≤ .993
Bourbon (1990)	.995 ≤ r ≤ .996
Bourbon et al. (1990)	
Experiment 1	.961 ≤ r ≤ .996
Experiment 2	.989 ≤ r ≤ .996
Experiment 3	.992 ≤ r ≤ .997
Bourbon (1996)	.971 ≤ r ≤ .998
Relative location in one dimension (Compensatory)	
Marken (2013)	r = .995
Powers (1989b)	.99841 ≤ r ≤ .99991
Relative location in two dimensions (Compensatory)	
Marken (1991)	r = .986
Matching distance between two pairs of horizontal lines	
Marken (1986)	
Experiment 1	.979 ≤ r ≤. .990
Experiment 2	.982 ≤ r ≤ .990
Experiment 3	r = .98

Note. 'Pursuit' refers to controlling the relationship of a cursor to target moving according to a randomly generated disturbance. 'Compensatory' refers to controlling the relationship of a cursor to a stationary target, despite disturbances applied to the cursor. Pearson *r* values represent the value, or range of values, for the correlations between the model behavior and the behavior of the individual measured across the time period of the task. They are reported to the decimal point value of the original paper.

The model-participants correlations in these replicated studies are described in Table 1. They are consistently above r =.94, across a total of 21 different experiments that have the same methodology but vary in a number of features, such as the number of participants and the number of spatial dimensions. This consistency is achieved across variants of the design despite the fact that the disturbances used to train the computer model – typically generated by a smoothed random sequence of numbers – were typically not the same as those used in the test phase. These findings were replicated in a recent study using root mean squared error as a measure of model fit rather than correlation coefficient (Parker, Tyson, Weightman, Abbott, Emsley, & Mansell, 2017). The average error rate at follow-up was 1.85%. Thus, the PCT model achieves a high level of accuracy in simulating the behavior of participants to ‘stimuli’ in the environment that they have not previously demonstrated. We would like to suggest that the model fits are particularly high for this group of studies not only because the theory may be valid, but because the *experimental methodology* used to test the theory fitted the criteria we are proposing to be essential to achieve replication, namely a closed-loop task collecting continuous data within an individual.

This visual tracking methodology has been extended to applied domains of human performance. A diverse range of experimental studies have been used to examine whether a particular controlled variable - *optical velocity* - is generally superior to a range of other possible controlled variables (e.g. optical acceleration), whether the object is thrown to oneself by another person (baseball catching), by oneself (throwing up a basketball and catching it), or moves autonomously (a toy helicopter) (Marken, 2005; Shaffer et al., 2013, 2015). The same basic model was replicated in each case. The design of these studies was critical to their robustness; they required data to be collected dynamically as part of an explicitly closed loop design, either within a computer simulation, or in a real-life task through coding movements recorded on video cameras as participants intercepted objects within their environment.

2. Evidence for applying the PCT methodology beyond visual tracking: proximity control

The next challenge has been to assess whether this experimental method generalizes to other controlled variables and disturbances, and to other contexts and samples. We will focus on a single controlled variable - proximity to others - because it can be widely studied across the behavioral sciences. For example, proximity control is a key feature of attachment during child development (Bowlby, 1972), grouping in animals (Niwa, 1994) and personal space in social and clinical psychology (Sundstrom & Altman, 1976). Yet it is important to note that, in principle, any traditional open-loop design can be reconceptualized as a closed-loop system when one takes into account the perceptual control abilities of the individual (Marken, 2013).

A series of studies involved frame-by-frame coding of videos of pairs of animals (rats and crickets) during food competition (Bell & Pellis, 2011; Bell, Judge, Johnson, Cade & Pellis, 2012). These researchers successfully used the TCV to show that the animals were controlling their proximity to one another. Following this, the researchers demonstrated that simulated agents controlling their proximity to one another showed qualitatively similar behavior to the animals (Bell, Bell, Schank & Pellis, 2015). Similarly, computational models of proximity control have been developed within social psychology to show qualitative similarities with the behavior of human crowds (McPhail et al., 1992). These studies clearly have the capacity to be extended to tests of computational models. For example, a paradigm to assess control of interpersonal distance has involved validating a computational model. Participants were recruited in groups of five, tested in each combination of pairs, and their distance from one another was recorded whilst standing for a conversation. The models were optimized to training data such that the preferred interpersonal distance for each individual (*reference value*) could be estimated, along with their *gain* for the control of this distance. This allowed the experimenter to use the model to predict the interpersonal distance of novel combinations of pairs of participants (Mansell, Rogers, Wood, & Marken, 2013).

A closed-loop design has also been used within clinical psychology as an alternative to paradigms that present a ‘threat stimulus’ and assess reaction time to a discrete response.

Participants use a joystick to control their continuous distance from a threatening image (a spider) within an image of a corridor on a computer screen (Healey, Mansell, & Tai, 2017). The participant needed to move the joystick continuously to adjust the distance because the spider moved of its own accord. One version of this paradigm reversed the relationship between joystick direction and distance from the image, in the same way as the reversal noted in the meta-analysis of flexion and extension studies described earlier (Laham, Kashima, Dix, & Wheeler, 2015). In both conditions, participants maintained a preferred distance, even though the actions required to achieve this were in the opposite direction between conditions.

Taking the literature above as a whole, it is clear that control of socially important and clinically relevant variables can be modeled accurately using valid closed-loop experimental designs. We now turn our attention to the circumstances when we might expect models (of any domain of the psychological sciences) to be less consistent and replicable, by considering further elements of perceptual control theory, namely the hierarchical organization of goals and reorganization.

3. Evidence for Generalizing PCT methodology to high level goals

The experiments we reviewed in the last section are required to have a very tight relationship between the IV (the experimental manipulation) and the DV (e.g. the behavioral response); in other words a very large effect size. The computational models of these tasks then emulate the task itself rather than various other processes involved in learning to perform them. It makes sense to build models of highly practised tasks before moving on to try to understand more sophisticated psychological processes, including the learning of novel and complex tasks. It also makes sense to model highly practised tasks from an ecological perspective - they are likely to be by far the most common activities people engage in - walking, talking, object manipulation, tracking moving objects, controlling distance - even though they may executed relatively automatically and outside awareness most of the time.

The nature of the experimental data to train computational models has been particularly highlighted by the work of Jeffrey Vancouver and colleagues. They have developed models of goal pursuit within organizational contexts (e.g. health services; education) drawing from PCT in order to attempt to provide a more accurate account than existing social learning theories (e.g. Bandura, 1978). Vancouver and colleagues have used study designs that fit their computational models: for example a task in which participants were given the role of a manager in a simulated work setting. The job was to allocate workers shift times to a schedule. This required them to balance time allocation and also meet the budget for the workforce (Vancouver, Putka, & Scherbaum, 2005). Their study serves as a worked example of the key stages of development when applying closed-loop designs to test a computational model in a novel domain of psychology.

Vancouver et al. (2005) helpfully broke down their design process into the following stages:

(1) *Circumscribe the phenomenon*. Any real world system is highly interconnected between individuals and their environments, and therefore attempting to model all systems would be unproductive (Forrester, 1968). Therefore, the researcher needs to specify the most important elements of the phenomenon of interest. Vancouver et al. (2005) studied the ‘goal level effect’ (Locke & Latham, 1990) and so they needed a paradigm in which a user could be assigned goals of different levels, apply decisions within the system, and get feedback on their effects.

(2) *Represent a specific context*. Next the researcher needs to select a relevant social context for the task. Vancouver et al. (2005) chose a simulation of computerized scheduling of nurses work patterns because it was relevant to an organizational context, and it allowed interactivity with the user within a closed loop.

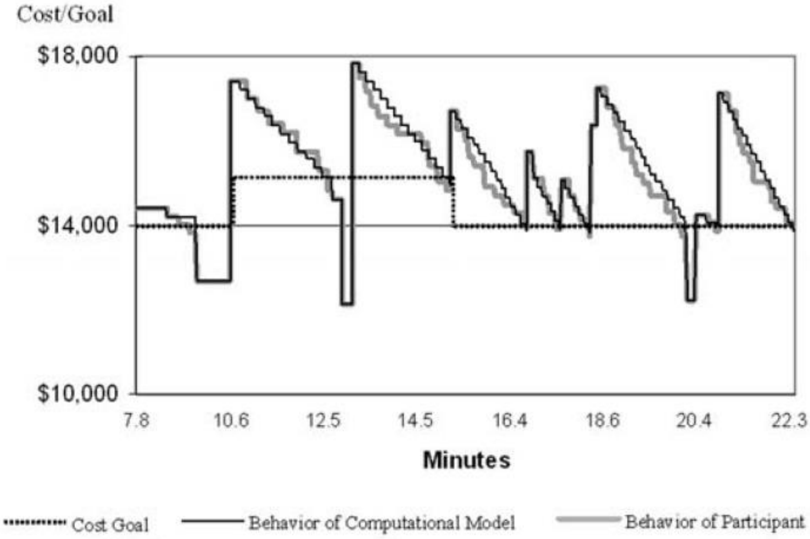
(3) *Select a software platform for computer modeling of the human participants*. Vancouver et al. (2005) describe a wide range of platforms, including spreadsheets (e.g. Marken, 1990), all purpose programming languages (e.g. C++, Basic), and simulation platforms, such as Vensim (www.vensim.com).

(4) *Construct the model*. The researcher uses the components of PCT to build a bespoke model that perceives the variables presented in the experimental task, compares these with reference values for the variables, and acts upon the task. Vancouver et al. (2005) produced a three-level hierarchy that implemented adjustments of the nurses schedules in order to manage its cost. This stage of the research clearly requires considerable thought and discussion as the model architecture is refined.

(5) *Test the model*. Vancouver et al. (2005) reported a number of test of their model. These included visual inspection of the graphs of the outputs of each participant and their computational model for direct comparison, and the calculation of intraclass correlations to estimate the model fit.

Figure 4 displays the example provided by Vancouver et al. (2005) and indicates the close match between participant and computer data. The computational model of performance of this task produced high correlations (all $r > .9$) between the individual participant behavior and that of the model for the majority of participants. Despite this being an abstract task, the level of fit was comparable to the target tracking tasks described above. Interestingly, it also revealed that an alternative model (based on another theory - Klein, 1989) applied to a minority of participants, thereby revealing the importance of studying individuals to gain a more nuanced understanding of individual differences, rather than attempting to generalize a single finding across groups. Taken together, they demonstrate that it is possible to follow the recommendations of our article when studying higher level, acquired goals. Vancouver and colleagues have also interweaved the modeling of discontinuous variables (e.g. the perception of events) into their models of continuous variables. We anticipate that such designs could be applied across social and cognitive psychology; for example an earlier study utilizing PCT showed that people counter trait words that are disturbances to their self concept (Robertson, Goldstein, Mermel & Musgrave, 1999). A future study could construct a computational model of this closed-loop process within individuals and test it against participant data. Next we turn to the acquisition of such goals during learning and development, and other higher-order processes such as reasoning and decision-making.

Figure 4. Plot of the relationship between time and cost on the scheduling task in one participant from Vancouver et al. (2005). The model shows the close match in performance of the computational model and the participant continuously during the task.



4. Generalizing PCT Models to Complex Psychological Processes

Designing studies to examine the learning and development of mental processes such as memory, reasoning and decision-making present a particular challenge to researchers (e.g. Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Kotseruba, Gonzalez, & Tsotsos, 2016). At present a large field of computational modeling is involved in simulating data from complex tasks that involve high levels of uncertainty both in terms of the participant's understanding of how to perform well, and in terms of the researchers' understanding of what the task is actually measuring. A characteristic example is the modeling of control performance in the classic dynamic systems task by Berry and Broadbent (1984). The task introduced a noise parameter to increase participant's uncertainty about the crucial features of performance. This meant that the data generated by both the participants and simulations showed highly variable behavior (e.g. Dienes & Fahey, 1995). In this situation, models can only be evaluated in terms of very broad qualitative comparisons between patterns of data across models and human behavior. In the sugar factory example, the learning rate in individual participants was very low, partly because noise was deliberately introduced into the procedure. However, even allowing for that, participants' performance was well below that of routine, highly practiced, tasks. The consequence of this approach to date for the sugar factory task is that the interpretation of computational models is judged according to broad qualitative impressions of the performance, across the sample, and rarely in terms of individual performance.

Within contemporary examples, the inclusion of random variability within computational models can permit individual differences and their correlates to be identified within group data (e.g. Guitart-Masip, Huys, Fuentemilla, Dayan, Duzel, & Dolan, 2012; Moutoussis, Dolan, & Dayan, 2016). Yet, it is when computational models begin to incorporate the processes within the individual that may be generating variability that model fit improves (Moutoussis et al., 2016). We will return to this point later when we describe the proposed role of intrinsically generated variability within a control architecture. Indeed, it is possible to produce models of the continuous data from individuals carrying out closed-loop tasks that do incorporate classically 'cognitive' processes (Vancouver & Purl, 2017).

Earlier we described the organization of a single control unit. The complete architecture describes how multiple control units are organized within a hierarchy (Powers et al., 1960; Powers, 1973). The outputs of each higher level in the hierarchy provide the reference values for the next layer down, enabling more complex variables to be controlled. This architecture also allows the operation of various *modes* of control as outputs can also form recurrent connections that are fed back upwards - for example allowing mental simulation of the perceptual results of action before they occur (Powers et al., 1960; Powers, 1973, 2008). This is thought to be the means through which we engage in mental imagery of our goals.

The model parameters of the work described above are stable; they do not change as the task progresses. As long as the output function of the model counters the disturbance, the perceptual variable is under control. However, more profound changes in environment can undermine control, for example where the *feedback function* of the environment is reversed. One example of this involved chicks on a treadmill that was set up so that when they ran faster they got closer to a cup of food. When this relationship was reversed experimentally – with the cup of food moving further away as they ran faster – they lost control because they consistently chased towards the food faster as it got further away (Hershberger, 1986). This study clearly supports the view that ‘approach behavior’ is not a triggered response, but part of a closed-loop process to try to control visual perception. Yet it also demonstrates how the basic control unit has limitations and so a new control strategy would need to be learned in order to restore control in some highly novel environments. For this reason, PCT incorporates a learning process that describes how new control systems are developed, optimized and conflict between them reduced - known as *reorganization* (Powers et al., 1960; Powers, 1973, 2008). Powers credited this process in part to earlier work by Ashby (1948) and Campbell (1960). This is the component of PCT that requires a stochastic process rather than a deterministic mechanism, and the element of the theory that involves a level of uncertainty. While the mathematical details of the algorithm are beyond the purpose of the current article, the key principle is that random variation in the parameters of a control unit is introduced in order to select the values that optimize control. As Powers (1989a) states, “The concept of a reorganizing system fills in a missing part of the control-system model: the explanation of how it got that way… how an organism could be organized, maintain its organization, and acquire new organizations that pertain specifically to the continued existence of the organism in a wide variety of changing environments.” (Powers, 1989a, p291).

Thus, from the perspective of PCT, we can make a hypothesis regarding when a computational model is likely to have a very high fit with human performance, and periods when it *cannot* have a tight fit, owing to the proposed self-generated randomness within the system. Specifically, we would expect accurate models to predominate in well-practised activities showing high levels of control by the participants. Periods of poorer fit would occur when control is poorer, until, through reorganization, the individual converges on the most adaptive perceptual variable to carry out the task effectively from their perspective, at which point model fit, and therefore replicability, should rise again suddenly, providing that the computational model controls the same variable as the human participant. Following this, both the human participant and the model should show a gradual improvement in control, and in model fit, as the parameters of the control system asymptotically approach their optimal values (Robertson & Glines, 1985).

There are a small number of PCT models that have designed experiments to test more complex elements of the PCT architecture. In terms of utilizing a hierarchy, one task involved participants controlling the relative distance of two independently moving objects on a computer screen (Marken, 1986). This closed-loop design fitted our stated criteria. A computational model requiring two hierarchical levels was constructed and optimized to the training data. It showed a high level of fit, similar to the position tracking studies described above.

A small number of studies have directly explored the stochastic learning process of reorganization using novel experimental designs (Pavloski, Barron, & Hogue, 1990; Robertson & Glines, 1985). For example, one study tested participants on a computerized 'game' that required them to detect and control three different perceptual variables varying in level of abstraction. These studies did not involve the construction of computational models. In a parallel line of work, computational models of the reorganization process have been developed and shown to have a qualitative resemblance to human performance (Marken & Powers, 1989; Powers, 2008), but these have not been tested against individual behavioral data.

Returning to the work of Vancouver and colleagues, they have further advanced the complexity of their models, including multiple goal pursuit (Vancouver, Weinhardt, & Schmidt, 2010), and learning and planning, drawing on further components of PCT in the process (Vancouver & Purl, 2017). These models show high internal validity in completing the tasks for which they were designed, although their match with behavior data has been more modest than the simpler models. This may be the case for two reasons. First, Vancouver and colleagues are striving to model complex processes in which multiple goals are modeled and ability to control is emerging, and so the models are likely to be less precise than the well-practised tracking and interception tasks typically modeled. Second, the existing research designs do not typically provide the data necessary to accurately test a computational model against individual, closed-loop data (Vancouver & Purl, 2017). Taken together, this body of work has made clear both the advantages and future challenges of experimental design for computational models based on PCT.

Future improvements in experimental designs for testing psychological theories

In this final section, we provide four further recommendations regarding how to further advance theory-testing through dynamic, closed-loop, assessments of individual performance.

1. Establish the individual specificity and stability of computational models and their parameters

First, we recognize a limitation in current testing of computational models. The test phases typically do not extend beyond the experimental session, immediately following the training phase for the model. Yet, psychological theories typically hypothesize both task-specific and enduring, person-specific traits or structures within individuals. Indeed, within the fields of social and personality psychology, the requirement for trait measures to have stability over time has long been upheld as an essential stage in establishing the validity of a new measure (Oppenheim, 1990). As we mentioned earlier, within psychophysics, it also seems feasible to establish consistency in perceptual parameters over time. Surely, it is no less important to establish test-retest reliability for computational models over longer periods than a single test session? Despite the importance of this, we have identified only two studies that tested the temporal stability of computational models (Bourbon et al., 1990; Bourbon, 1996), and none that directly assessed whether models were individual-specific. In our lab, we have recently completed a study of participants carrying out a one-dimensional tracking task and showed that the computational models of individuals derived from PCT were specific to each individual, and highly stable over one week (Parker et al., 2017).

Going beyond the above work, one might argue that a fully robust test of a computational model of an individual will have the capacity to model performance that is based on both task-specific (and therefore varies over time within an individual depending on their current task) and task-invariant parameters and strategies. In other words, we come even closer to testing the validity of a theory in modeling an individual when models and their parameters have consistencies across different tasks. We know of one study achieving this -

although not within a closed-loop task - terming the methodology ‘cognitive tomography’ (Houlsby, Huszar, Ghassemi, Orban, Wolpert, & Lengyel, 2013). In this study, computational models of individual participants were constructed based on their judgment of familiarity of an array of facial images. The models utilized idiosyncratic internal representations of facial images to successfully complete the original task, and the same models fitted individual performance during a subsequent, different task that involved identifying ‘the odd one out‘ in an array of facial images. Future work on models of closed-loop activities can take inspiration from this design. They would examine the degree of consistency in models and their parameters on *different* tasks that are nonetheless proposed to require the control of the same perceptual variables.

2. Carry out comparative tests between computational models against behavioral data

An appropriate methodology for computational model testing against behavioral data is to compare and contrast the fit of competing models (Guitart-Masip et al., 2012; Moutoussis et al., 2016; Vancouver et al., 2010). A small number of studies have compared a PCT model against a ‘cognitive’ or ‘open-loop’ model, and found that the PCT model to be superior in fit to behavioral data (Bourbon & Powers, 1999; Marken, 2013). However, in both cases, the competing model was constructed by experts in the PCT models. In the future, collaborative work is required so that competing models derived from distinct theories are contrasted in their degree of fit with the individual data of specific tasks. Particular attention will need to be paid to ensure that competing models have equal capacities to optimize their parameters to the individual for a fair comparison.

3. Build and test physical simulations and robotic devices against behavioral data

There is an increasing appreciation that robotics provide a testing ground for psychological theory (Sleek, 2016). Yet, at present the aim is typically to produce qualitatively similar behavior to that of a human, rather than to attempt to quantitatively match the behavior of a human carrying out the same task. Where a robotic device is required to implement a computational model, the test becomes increasingly challenging. For example, the typical model of tracking performance permits the computer to move the cursor on the screen directly through the electronics of the computer. The model is not required to move a joystick or a mouse like the human user. For this reason, the functional architecture required to *embody* movement is not typically modeled; the model specification is incomplete. Whilst it could be argued that modeling anatomy is an overly complex requirement for a psychology experiment, there is clear evidence that body arrangement and posture are significantly associated with measures of cognition and emotion (Niedenthal, Barsalou, Winkielman, Krauth-Gruber & Ric, 2005). Indeed, the body itself may have a key role in the process of control and computation (Füchslin, Dzyakanchuk, Flumini, Hauser, Hunt, Luchsinger et al., 2013). Furthermore, modeling anatomy adds additional complexity to one’s assumptions regarding the design a computational model because the different joints within a limb act as self-generated disturbances to one another, which need to be counteracted in an ongoing manner (Powers, 2008).

Despite the above challenges, the potential leap further towards robotic technology has intermediate steps that make it somewhat simpler. For example, computer simulations can be used within computational models to emulate the physics of the environment (e.g. Kennaway, 1999; Powers, 1999). However, such approaches are open to criticism regarding the accuracy and oversimplification of such modeling of the environment. An alternative step in this direction that our lab has piloted is the use of force-feedback devices. They allow a computational model to demonstrate its ability to use physical movement in the real world (the computer-generated motion of a joystick) to close the loop between an output signal and a shift in the variable that is being controlled (the location of the cursor). A recent effective test of the PCT architecture has involved embodying it within a rover vehicle (Young, 2017).

Yet, arguably only robotic structures that are more analogous to human anatomy will permit a direct test of an embodied model (e.g. Guizzo & Ackerman, 2012). A shift of this kind towards artificial life as a means to test psychological theory has its own philosophical and ethical challenges (Jordan, 2008), but need not be seen to be outside the realms of psychological science.

How can researchers make the transition we are recommending?

We recognize the huge challenges involved for researchers to shift from existing conceptualizations of behavior and their associated methodologies, to the alternative we are suggesting. The inherent difficulties of any cultural, organizational or scientific system to make changes are well recognised (Kuhn, 1962; Todnem By, 2005). There are two sides to any system confronting change.

On one side are the perceived benefits of keeping the status quo alongside the perceived fears and costs of shifting to a new set of conceptual and methodological assumptions. It is clearly easier to stick with the existing system because it is supported by the vast majority of textbooks, school, college and degree courses, and academic journals. There is a realistic fear that, at least in the short term, producing research based on a different set of assumptions may not be published, nor supported by funding bodies.

On the other side are the costs of keeping the existing system and the benefits of the shift. We have tried to make a clear case that the cost of keeping the existing system will be no resolution of the replication crisis, which in turn will limit scientific progress. In addition we have described evidence for the benefits of more accurate and consistent replication of research findings. There are several other reasons that researchers considering making the shift should be heartened:

1) *Resources.* The methods to examine perceptual control are supported by a variety of resources. We summarized the TCV, which is also described in a number of books with a range of illustrative examples (e.g. Marken, 2014; Runkel, 2007). Software is also freely available to collect data on visual tracking, and to build models of crowd behavior (Powers, 2008). We have also summarized a number of tools available to build new computational models, including spreadsheets (Marken, 1990), and model-building software (e.g. Vensim).

2) *Contemporary technology.* The perceptual control paradigm lends itself easily to a number of contemporary advances in research technology that are already popular with researchers yet appear to have not yet reached their full potential. For example, virtual reality systems have the capacity to collect continuous real-time data on bodily movements, and to present continuous information to the visual and auditory modalities. Yet, this data is rarely used to train and test computational models of the psychological processes involved. A similar case can be made for smartphones and advanced robots.

3) *Coherent framework.* The capacity for the integration of research findings is enhanced by this approach. Through considering behavior as the control of perceptions, research on behavior across a variety of contexts can be seen to share the same underlying patterns. One example is within the field of psychopathology. Increasingly it is becoming recognized that the wide variety of symptoms of different psychiatric disorders share a conceptual structure; they reflect attempts to control various emotional experiences in ways that limit the individual's opportunity to process and resolve longstanding goal conflict (Mansell, Carey, & Tai, 2015). This conceptualization in turn can inform a more efficient and universal psychological intervention (Alsawy, Mansell, McEvoy & Tai, 2015).

In sum, despite the costs of change, there are appropriate means available to make the change and a range of benefits to be identified in the long term.

Conclusions

In the future, we hope that increasingly larger numbers of researchers from diverse backgrounds will recognize and use perceptual control methodology. This will allow the field of psychological science to continue to move from null hypothesis testing of aggregated groups of participants designs towards using models to prototype individual performance as a robust test of the 'truth' of a scientific theory. We have proposed that such a scientific venture will only succeed if close attention is paid to the experimental design used to collect participant data. We also showed examples of where these advances in research design advance scientific knowledge and test psychological theories even in the absence of a computer model.

Specifically, we made the recommendations that studies collect continuous individual data from participants carrying out activities that are 'closed loop' in nature and that ideally generate very high effect sizes for the IV-DV relationship. We offered research on perceptual control theory as an illustration of where these requirements have been met and led to a number of accurate models of real world psychological phenomena. We have recognized that there is however a tension between the complexity of computational models, and the advances in experimental design required to test them robustly at the individual level. We see this as an inevitable feature of progress in this area and anticipate that adhering to the recommendations for experimental design we have described could act as a helpful rubric to hone research development in this field.

At its pinnacle, researchers from all backgrounds will be able to directly compare the capacity for different (and potentially theoretically distinct) computational models of individuals to match task performance. From here, replication can proceed. The researcher attempts to replicate the degree of fit between a computational model derived from a theory and continuous data from individuals. Ultimately, a valid theory should permit replication of a model not only within the same participant over time, but across different tasks to the degree that they share the same components. We have specified the conditions under which near exact replication of individual models is feasible (i.e. during highly practiced tasks in which the participant manages high levels of control) and those in which it will retain some level of 'noise' but nonetheless remain a robust test of psychological theory (i.e. during the learning and practice of more difficult versions of the same task, or novel tasks that require random changes to converge on an effective control strategy). We aim to pursue all of these considerations within our future work on 'replication science'.

Acknowledgements

Thank you to Andrew Willett for compiling the list of studies for Table 1 and to Alex Gomez-Marin, Maximilian Parker, Michael Moutoussis and earlier anonymous reviewers for their generous feedback on earlier drafts and additional contributions.

Footnotes

1.It is tempting to think that a number of contemporary approaches to computational modeling might occupy the same ground as PCT. Most prominent here are predictive coding accounts including those that involve the free energy principle or active inference (e.g. Seth, Suzuki, & Critchley, 2012). The contemporary view is that the fundamental role of the brain is to make predictions about the environment through its own internal models to determine its behavioral response. In contrast, according to PCT, the control of perceptual input is the fundamental role of the brain and not prediction. Prediction can occur as part of the process of perceptual control but it is not a necessary requirement for psychological functioning. Control requires only the establishment of adaptive perceptual goals, many of which are described in the current article (e.g. Bell, 2014; Marken, 2005). Indeed, some of the most contemporary empirical work putting to the test established models of the brain as a 'prediction machine' actually support the view that sensory input is a controlled goal-state rather than a predicted outcome (e.g. Niziolek, Nagarajan, & Houde, 2013). Nonetheless, we expect that both accounts have strengths and limitations. Rather than attempting to integrate or contrast PCT with contemporary prediction error accounts in the current article, we have chosen to simplify our message by focusing on the specific contributions of a PCT account to research methodology.

References

Alsawy, S., Mansell, W., Carey, T. A., McEvoy, P., & Tai, S. (2014). Science and practice of transdiagnostic CBT: A Perceptual Control Theory (PCT) approach. *International Journal of Cognitive Therapy, 7(4),* 334-359.

Anderson, J.R. & Bower, G.H. (1974). A propositional theory of recognition memory. *Memory and Cognition, 2,* 406 - 412.

Ashby, W. R. (1948). Design for a Brain. *Electronic Engineering*, *20,* 379–383.

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist, 33(4),* 344-358.

Baum, W. M., Reese, H. W., & Powers, W. T. (1973). Behaviorism and feedback control. Science, 181, 1114.

Bell, H. C. (2014). Behavioral variability in the service of constancy. *International Journal of Comparative Psychology, 27(2),* 196-217.

Bell, H. C., & Pellis, S. M. (2011). A cybernetic perspective on food protection in rats: simple rules can generate complex and adaptable behaviour. *Animal Behaviour*, *82* (4), 659-666.

Bell, H. C., Judge, K. A., Johnson, E. A., Cade, W. H., & Pellis, S. M. (2012). How is a cricket like a rat? Insights from the application of cybernetics to evasive food protective behaviour. *Animal Behaviour, 84(4),* 843-851.

Bell, H. C., Bell, G. D., Schank, J. A., & Pellis, S. M. (2015). Evolving the tactics of play fighting: insights from simulating the "keep away game" in rats. *Adaptive Behavior, 23(6),* 371-380.

Berry, D.C. & Broadbent, D.E. (1984). On the relationship between task performance and associated verbalizable knowledge, *The Quarterly Journal of Experimental Psychology Section A, 36:2,* 209-231.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113(4),* 700-765.

Bourbon, W. T. (1989). A control-theory analysis of interference during social tracking. In W. A. Hershberger (Ed.), *Volitional Action: Conation and Control* (pp. 235-251). North-Holland, Amsterdam: Elsevier Science Publishers.

Bourbon, W. T. (1990). Invitation to the dance: Explaining the variance when control systems interact. *American Behavioral Scientist*, 34 (1), 95-105.

Bourbon, W. T. (1996). On the accuracy and reliability of predictions by perceptual control theory: Five years later. *The Psychological Record,* 46, 39-47.

Bourbon, W. T., & Powers, W. T. (1999). Models and their worlds. *International Journal of Human-Computer Studies, 50(6),* 445-461.

Bowlby, J. (1972). *Attachment: Attachment and loss. Vol 1.* Penguin Books.

Blyth, C. R. (1972). On Simpson's paradox and the sure-thing principle. Journal of the *American Statistical Association, 67(338),* 364-366.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14(5),* 365-376.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review, 67(6),* 380-400.

Carey, T. A., Tai, S. J., Mansell, W., Huddy, V., Griffith, R., & Marken, R. S. (2017). Improving professional psychological practice through an increased repertoire of research methodologies: Illustrated by the development of MOL. *Professional Psychology: Research and Practice, 48(3),* 175-182.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review, 97(3),* 332-361.

Colquitt, J. A., Scott, B. A., Judge, T. A., & Shaw, J. C. (2006). Justice and personality: Using integrative theories to derive moderators of justice effects. *Organizational Behavior and Human Decision Processes, 100(1),* 110-127.

da Silva, J. A., & Fukusima, S. S. (1986). Stability of individual psychophysical functions for perceived distance in natural indoor and outdoor settings. *Perceptual and Motor Skills, 63(2),* 891-902.

Dienes Z., & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 848-62.

El Hady, A. (2016). *Closed Loop Neuroscience.* Academic Press.

Forrester, J. W. (1968). *Principles of systems.* Cambridge, MA: MIT Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews. Neuroscience, 11(2),* 127-138.

Fum, D., Del Missier, F., & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research, 8*, 135-142.

Füchslin, R. M., Dzyakanchuk, A., Flumini, D., Hauser, H., Hunt, K. J., Luchsinger, R. H., ... & Walker, R. (2013). Morphological computation and morphological control: steps toward a formal theory and applications. *Artificial Life, 19(1),* 9-34.

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America, 101(36),* 13124-13131.

Glautier, S. (2013). Revisiting the learning curve (once again). *Frontiers in psychology, 4*, 982.

Gluth, S. & Rieskamp, J. (2017) Variability in behavior that cognitive models do not explain can be linked to neuroimaging data. *Journal of Mathematical Psychology*.

Glover, S. (2004). Separate visual representations and the planning and control of action. *Behavioural and Brain Sciences*, 27, 3-78.

Guitart-Masip, M., Huys, Q. J., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage, 62(1),* 154-166.

Guizzo E., & Ackerman, E. (2012). How rethink robotics built its new Baxter robot worker. *IEEE Spectrum: September.*

Hayes, A. M., Laurenceau, J. P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review, 27(6),* 715-723.

Healey, A., Mansell, W., & Tai, S. (2017). An experimental test of the role of control in spider fear. *Journal of Anxiety Disorders, 49,* 12-20.

Heisenberg, M., & Wolf, R. (1992). The sensory-motor link in motion-dependent flight control of flies. *Reviews of Oculomotor Research, 5,* 265-283.

Henriques, D. Y., Klier, E. M., Smith, M. A., Lowy, D., & Crawford, J. D. (1998). Gaze-centered remapping of remembered visual space in an open-loop pointing task. *The Journal of Neuroscience, 18(4),* 1583-1594.

Hershberger, W. A. (1986). An approach through the looking-glass. *Animal Learning & Behavior, 14(4),* 443-451.

Horton, R. (2015) Offline: What is medicine's 5 sigma? *Lancet, 385*, 1380.

Houlsby, N. M., Huszar, F., Ghassemi, M. M., Orban, G., Wolpert, D. M., & Lengyel, M. (2013). Cognitive tomography reveals complex, task-independent mental representations. *Current Biology, 23 (21),* 2169-2175.

Hulin, C. L., & Ilgen, D. R. (2000). *Computational modeling of behavior in organizations: The third scientific discipline.* Washington, DC: American Psychological Association.

Iaria G., Petrides, M., Dagher, A., Pike, B., & Bohbot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *Journal of Neuroscience, 23, 5945–5952.*

Ioannidis, J. P. (2005) Why most published research findings are false. *PLoS Medicine, 2.* DOI: 10.1371/journal.pmed.0020124

Jordan, J. S. (2008). Wild agency: nested intentionalities in cognitive neuroscience and archaeology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363(1499),* 1981-1991.

Katz, D., & Kahn, R. L. (1966). *The social psychology of organizations.* New York: John Wiley.

Kazdin, A. E. (2011). *Single-case Research Designs: Methods for Clinical and Applied Settings*. Oxford University Press.

Kennaway J. (1999). Control of a multi-legged robot based on hierarchical perceptual control theory. *Journal of Perceptual Control Theory, 1,* 1–7.

Klein, H. J. (1989). An integrated control theory model of work motivation. *Academy of Management Review, 14(2),* 150-172.

Kotseruba, I., Gonzalez, O. J. A., & Tsotsos, J. K. (2016). A review of 40 years of cognitive architecture research: Focus on perception, attention, learning and applications. arXiv preprint arXiv:1610.08602.

Kuhn, T.S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Laham, S. M., Kashima, Y., Dix, J., & Wheeler, M. (2015). A meta-analysis of the facilitation of arm flexion and extension movements as a function of stimulus valence. *Cognition and Emotion, 29,* 1069-1090.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society B: Biological Sciences, 352(1358),* 1231-1239.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal-setting and task performance.* Englewood Cliffs, NJ: Prentice Hall.

Logue, A. W. (1976). Individual differences in magnitude estimation of loudness. *Perception & Psychophysics, 19*(3), 279-280.

Mansell, W., Carey, T. A., & Tai, S. J. (2015). Classification of psychopathology and unifying theory: The ingredients of a Darwinian paradigm shift in research methodology. *Psychopathology Review, 2,* 129-153.

Marken, R. S. (1997) The dancer and the dance: methods in the study of living control systems. *Psychological Methods*, *2(4),* 436-446.

Marken, R. S. (1986). Perceptual organization of behavior: A hierarchical control model of coordinate action. *Journal of Experimental Psychology: Human Perception and Performance, 12(3),* 267-276.

Marken, R. S. (1991). Degrees of freedom in behavior. *Psychological Science,* 2 (2), 92-100.

Marken, R. S. (2005) Optical trajectories and the informational basis of fly ball catching, *Journal of Experimental Psychology: Human Perception and Performance, 31,* 630–634.

Marken, R. S. (2013). Taking purpose into account in experimental psychology: Testing for controlled variables. *Psychological Reports, 112,* 184-201.

Marken, R. S. (2014). *Doing Research on Purpose: A Control Theory Approach to Experimental Psychology*. New View Publications.

Marken, R. S., & Carey, T. A. (2015). Understanding the change process involved in solving psychological problems: A model ‐ based approach to understanding how psychotherapy works. *Clinical Psychology & Psychotherapy, 22(6),* 580-590.

Mansell, W., Carey, T. A., & Tai, S. J. (2015). Classification of psychopathology and unifying theory: The ingredients of a Darwinian paradigm shift in research methodology. *Psychopathology Review, 2, 129-153.* DOI: 10.5127/pr.036114

Mansell, W., Rogers, D., Wood, K. & Marken, R.S. (2013). Is behaviour the control of perception? A computer model of personal space during a dyadic interaction. *Paper presented at the International Ethology Conference, August 2013, Newcastle, UK.*

Marken, R. S., & Powers, W. T. (1989). Levels of Intention in Behavior. In W. A. Hershberger (Ed.), *Volitional Action: Conation and Control* (pp. 409-430).

Marken, R. S. (1990). Spreadsheet analysis of a hierarchical control system model of behavior. *Behavior Research Methods, Instruments, & Computers, 22(4),* 349-359.

McClelland, K. (2014). Cycles of conflict: A computational modeling alternative to Collins's theory of conflict escalation. *Sociological Theory, 32(2)*, 100-127.

McPhail, C., Powers, W. T., & Tucker, C. W. (1992). Simulating individual and collective action in temporary gatherings. *Social Science Computer Review, 10(1),* 1-28.

Moutoussis, M., Dolan, R. J., & Dayan, P. (2016). How people use social information to find out what to want in the paradigmatic case of tnter-temporal preferences. *PLoS Computational Biology, 12(7),* e1004965.

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). Bayesian inferences about the self (and others): A review. *Consciousness and Cognition, 25,* 67-76.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44(1),* 190-204.

Nicholson, L., Adams, R., & Maher, C. (2000). Magnitude estimation of manually assessed elastic stiffness: stability of the exponent. Perceptual and Motor Skills, 91(2), 581-592.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9(3),* 184-211.

Niwa, H. S. (1994). Self-organizing dynamic model of fish schooling. *Journal of Theoretical Biology, 171(2),* 123-136.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience, 33(41),* 16110-16116.

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people. *Frontiers in Psychology, 7.* http://dx.doi.org/10.3389/fpsyg.2016.00934

Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science 349*, DOI: 10.1126/science.aac4716

Parker, M., Tyson, S., Weightman, A., Abbott, B., Emsley, R., & Mansell, W. (2017). Perceptual control models of pursuit manual tracking demonstrate individual specificity and parameter consistency. *Manuscript submitted for consideration.*

Pashler, H., & Wagenmakers, E.-J., 2012. Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science 7(6),* 528-530.

Pavloski, R.P., Barron, G.T., Hogue, M.A. (1990). Reorganization: learning and attention in a hierarchy of control systems. *American Behavioral Scientist 34(1),* 32–54.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109(3),* 472-491.

Powers, W. T. (1973). *Behavior: The control of perception.* New York, USA: Hawthorne.

Powers, W. T. (1978) Quantitative Analysis of Purposive Systems: Some Spadework at the Foundations of Scientific Psychology. *Psychological Review,* 85, 5, 417-435.

Powers, W. T. (1979a). The nature of robots. Part 1: Defining behavior. *Byte, 4(6),* 132-144.

Powers, W. T. (1979b). The nature of robots: Part 2. simulated control system. *Byte, 4(7),* 144-152.

Powers, W. T. (1979c). The nature of robots. Part 3: A closer look at human behavior. *Byte, 4(8),* 94-116.

Powers, W. T. (1979d). The nature of robots: Part 4. Looking for controlled variables. *Byte, 4(9)*, 96-112.

Powers, W. T. (1989a). *Living Control Systems.* The Control Systems Group, Inc. Gravel Switch: KY.

Powers, W. T. (1989b). Quantitative Measurements of Volition: A pilot-study. In W. A. Hershberge (Ed.), *Volitional Action: Conation and Control* (pp. 315-332) North-Holland, Amsterdam: Elsevier Science Publishers.

Powers, W. T. (1990). Control theory and statistical generalizations. *The American Behavioral Scientist, 34(1), 24*.

Powers, W. T. (1992). *Living Control Systems II: Selected papers of William T. Powers.* The Control Systems Group, Gravel Switch, Kentucky, USA.

Powers, W. T. (1999). A model of kinesthetically and visually controlled arm movement. *International journal of human-computer studies, 50(6),* 463-479.

Powers, W. T. (2005). *Behavior: The control of perception.* New Canaan, CT, USA: Benchmark.

Powers, W. T. (2008). *Living Control Systems III: The fact of control.* New Canaan, CT, USA: Benchmark Publications Inc. Software available at www.billpct.org.

Powers, W. T., Clark, R. K., and McFarland, R. L. (1960a). A general feedback theory of human behavior. Part I. *Perceptual and Motor Skills, 11,* 71-88.

Powers, W. T., Clark, R. K., and McFarland, R. L. (1960b). A general feedback theory of human behavior. Part II. *Perceptual and Motor Skills, 11,* 309-323.

Robertson, R. J., & Glines, L. A. (1985). The phantom plateau returns. *Perceptual and Motor Skills, 61(1),* 55-64.

Robertson, R. J., Goldstein, D. M., Mermel, M., & Musgrave, M. (1999). Testing the self as a control system: Theoretical and methodological issues. *International Journal of Human-Computer Studies, 50(6),* 571-580.

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: a quiet methodological revolution. *The American Psychologist, 65(1),* 1-12.

Runkel, P. J. (2007). *Casting nets and testing specimens: Two grand methods of psychology.* Living Control Systems Publishing.

Runkel, P. J., & McGrath, J. E. (1972). *Studying human behavior.* New York: Holt.

Schlesinger, M., & Parisi, D. (2001). The agent-based approach: A new direction for computational models of development. *Developmental Review, 21(1),* 121-146.

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology, 2,* 395.

Shaffer, D. M., Marken, R. S., Dolgov, I., & Maynor, A. B. (2013). Chasin'choppers: Using unpredictable trajectories to test theories of object interception. *Attention, Perception, & Psychophysics, 75(7),* 1496-1506.

Shaffer, D. M., Marken, R. S., Dolgov, I., & Maynor, A. B. (2015). Catching objects thrown to oneself: Testing control strategies for object interception in a novel domain. *Perception, 44(4),* 400-409.

Shoda, Y., LeeTiernan, S. , & Mischel, W. (2002). Personality as dynamical system: Emergence of stability and distinctiveness from intra-and interpersonal interactions. *Personality and Social Psychology Review, 6(4),* 316-325.

Sleek, S. (2016). Robots for research. *Observer, 29(8).*

Smith, E. R., & Conrey, F. R. (2007). Agent-based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review, 11*(1), 87-104.

Stevens, S.S. (1957). On the psychophysical law. *Psychological Review. 64*(3), 153–181. doi:10.1037/h0046162.

Stevens, J.R. (2017). Replicability and reproducibility in comparative psychology. *Frontiers in Psychology, 8,* 862. doi: 10.3389/fpsyg.2017.00862

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18(6),* 643-662.

Sundstrom, E., & Altman, I. (1976). Interpersonal relationships and personal space: Research review and theoretical model. *Human Ecology, 4(1),* 47-67.

Taylor, F., & Birmingham, H. (1948). Studies of tracking behavior; the acceleration pattern of quick manual corrective responses. *Journal of Experimental Psychology, 38(6),* 783-795.

Teghtsoonian, M., & Teghtsoonian, R. (1983). Consistency of individual exponents in cross-modal matching. *Attention, Perception, & Psychophysics, 33(3),* 203-214.

Todnem By, R. (2005). Organizational change management: A critical review. *Journal of Change Management, 5(4),* 369-380.

Tolin, D. F., Lohr, J. M., Lee, T. C., & Sawchuk, C. N. (1999). Visual avoidance in specific phobia. *Behaviour Research and Therapy, 37(1),* 63-70.

Vancouver, J. B., & Purl, J. D. (2017). A computational model of self-efficacy's various effects on performance: Moving the debate forward. *The Journal of Applied Psychology.*

Vancouver, J. B., Putka, D. J., & Scherbaum, C. A. (2005). Testing a computational model of the goal-level effect: An example of a neglected methodology. *Organizational Research Methods, 8(1),* 100-127.

Vancouver, J. B., & Weinhardt, J. M. (2012). Modeling the mind and the milieu computational modeling for micro-level organizational researchers. *Organizational Research Methods, 15(*4), 602-623.

Vancouver, J. B., Weinhardt, J. M., & Schmidt, A. M. (2010). A formal, computational theory of multiple-goal pursuit: integrating goal-choice and goal-striving processes. *The Journal of Applied Psychology, 95(6),* 985-1008.

Warmerdam, L., van Straten, A., Jongsma, J., Twisk, J., & Cuijpers, P. (2010). Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: Exploring mechanisms of change. *Journal of Behavior Therapy and Experimental Psychiatry, 41(1),* 64-70.

Wiener, N. (1948). *Cybernetics*. Paris: Hermann.
Young, R. (2017). A general architecture for robotics systems: A perception-based approach to artificial life. *Artificial Life, 23(2),* 236-286. doi: 10.1162/ARTL_a_00229