CURRENT ISSUES – PERSPECTIVES AND REVIEWS

The Limits of Artificial Stimuli in Behavioral Research: The Umwelt Gamble

David C. Lahti

Department of Biology, Queens College, City University of New York, New York, NY, USA

(Invited Review)

Correspondence
David C. Lahti, Department of Biology, Queens College, City University of New York, 65-30 Kissena Blvd., Flushing, NY 11367, USA.
E-mail: david.lahti@qc.cuny.edu

Received: November 11, 2014
Initial acceptance: December 9, 2014
Final acceptance: December 31, 2014
(S. Foster)

doi: 10.1111/eth.12361

Keywords: umwelt, control, experimental design, artifact, eggs, brood parasitism

Abstract
The use of artificial stimuli in behavioral experimentation is pervasive and well preceded. A perspective by Hauber et al. (2014) Ethology describes advantages of this approach and highlights the use of model bird eggs and artificial egg coloration in research on egg rejection in the context of brood parasitism. Here, as a companion piece focused on quality control, I outline limitations and pitfalls associated with the use of artificial stimuli. In general, the practice makes assumptions about the perception and cognition of the study organism, therefore involving what could be called an umwelt gamble. The magnitude of this gamble and the prospects for interpretable results depend on the experimental design, the particular stimulus, and its intended role. Common roles are (1) as a representative stimulus to be generalized to a broader class; (2) as a substitute for a natural stimulus; (3) as a modification or exaggeration of natural stimuli; or (4) as an entirely novel stimulus. Whether the gamble is successful—whether the methodology navigates peculiarities of the study organism in the way the researchers intend—can be tested with controls that function as artifact detection tests. Given the propensity of animals to be biased, sometimes in unforeseen ways, in the way they perceive and interpret their environments, researchers should be careful when considering the use of artificial stimuli, weighing the advantages against the risks in any particular case.

As a new graduate student, in preparation for a pilot study of egg rejection in the African village weaverbird (Ploceus cucullatus), I and two fellow researchers meticulously sanded wooden eggs to size and shape, filled them with metal filings until they were just the right mass according to the oology books, and painted them three specific colors that we had painstakingly mixed until they, to our eyes, were within the range of egg colors we found in museum collections of that species. Later, in The Gambia, I discovered that all of this care was taken for naught. The birds, when they sought to reject the wooden egg, ended up jamming it deep into the lining of the nest, very often breaking their own eggs in the process. My treasured artifacts had elicited (1) a novel rejection method for the species (burial), (2) enormous costs of rejection, and (3) frequent nest desertion—all without a shred of biological relevance. My mistake, as binoculars would later inform me, was that I had failed to anticipate that the village weaver would often puncture eggs rather than grasp them in the process of removal and, of course, could not puncture the wood (Victoria 1972; Prather et al. 2007). I would also soon learn of my equally significant failure to appreciate the difference between the bird’s visual system and my own (e.g., Bennett & Théry 2007; Stevens 2011). Consequently, my experimental objects had also elicited (4) an egg rejection rate that included some unknown level of bias due to my use of color charts for comparison and to the paint’s unrealistic texture and lack of reflectance in the ultraviolet (UV) wavelength range. The only result I was able to take back to the bungalow was the
realization that using artificial stimuli involves a sort of wager regarding the capabilities and tendencies of the study organism, from perception, to cognition, to behavior.

The use of artificial stimuli in behavioral experimentation is pervasive and rich in precedent, yet controversial. Few would dispute that many interesting and biologically relevant results have proceeded from their use. Tinbergen’s (1951) description of graylag goose (*Anser anser*) retrieval of golf balls and door knobs as although they were eggs revealed the generality of fixed action patterns despite the artificiality and unfamiliarity of these objects to the bird—in fact, precisely because of their unfamiliarity. Hauber et al. (2014) highlight recent work that likewise features bird eggs, especially in the context of brood parasitism research. Indeed, a Web of Science literature search of experimental studies dealing with egg rejection in the context of avian brood parasitism yielded 54 studies between 2009 and 2014 that employ artificial eggs or eggs artificially colored (Fig. 1, and Appendix). The practice appears to be increasing in popularity. Hauber et al. (2014), some of whose authors also coauthor 20 of those studies, endorse the use of artificial stimuli and stress the continuity between modern avian brood parasitism research and classic ethology in this respect. The main point of their paper, however, is to alert us to and argue against a perceived knee-jerk reaction in peer review, whereby artificial stimuli are denigrated as biologically irrelevant. They accomplish this mainly by outlining the potential benefits of such stimuli, benefits that might be more difficult or even impossible to achieve through natural stimuli alone. One question, however, they do not address: What is it that distinguishes a good, biologically relevant and scientifically productive use of artificial stimuli, from one that demonstrates nothing but an idiosyncratic behavioral response to an oddity? I will attempt to address that question here.

The central challenge with artificial stimuli is to maintain a strong relationship with the real world. When a researcher steps out of the natural to explain the natural, this creates a significant responsibility to establish a thread of connection to nature and eventually follow it back. Criticisms can be fierce both at the point of departure and that of re-entry. The mere presentation of an artificial stimulus can seem to break all connection with nature and so be useless; and, subsequently, even if the experiment works the interpretation can be seen as limited only to those outlandish objects and nothing more. The researchers, in turn, might respond that without their artificial stimuli the questions they ask cannot be addressed. The main point of this paper is that it is the researchers using artificial stimuli who have the burden of proof, because they have taken a particular kind of gamble that is subject to failure owing to our ignorance of the perceptual and cognitive architecture of our study animals. The critics have a right to be conservative, until they are presented with evidence that their doubts are misplaced.

The ethological concept of the *umwelt* is, literally, the ‘world unto’ the organism—the world as the animal perceives it (van Uexküll 1909). One of the most rewarding experiences for an instructor in an introductory animal behavior class is to convey the diversity in the ways animals experience and interpret their surroundings. We can never fully know what it is like to be a bat (Nagel 1974), because of our inability to echolocate, to hear above 20 kHz, and so on. The authors of Hauber et al. (2014) have been instrumental in disseminating similar realizations about the bird *umwelt*—we cannot fully know what it is like to be a bird because, among other things, many of them see in the UV, they have four types of visual cones rather than three, and they process visual and auditory information much more quickly than we do. Thus, when we perform experiments that assume continuity between our own perceptual and cognitive

![Fig. 1: Recent experimental studies of egg rejection in the context of avian brood parasitism that employ artificial egg stimuli.](image-url)
substitute for the two kinds of stimuli do not differ in ways that undermine the substitution. Hence the widespread use of conspecific eggs to test egg recognition even when the goal is explicitly related to cuckoo parasitism. This is assumed to be justifiable in cases where cuckoos mimic host eggs, for instance, especially when their sizes and shapes overlap (e.g., Lahti 2006). However, many assumptions that subjects cannot tell the difference—or, more formally, as long as behavioral responses to the two kinds of stimuli do not differ in ways that undermine the substitution. The extent to which a single stimulus can represent an entire class is always debatable, because stimuli vary within a class, and a researcher cannot be certain that the feature of interest is responsible for the subject’s response rather than some other feature. For this reason, for instance, the bird song community now generally recognizes that playback studies must employ multiple experimental song stimuli (McGregor et al. 1992). In general, as the sample of stimuli grows, the sample becomes increasingly representative of the class, and the probability decreases that some unknown feature is responsible for the response. Applying this general rule to artificial stimuli, researchers should create not one but several stimuli for use in experiments, to average across features that are not intended to be salient. The partial exception to this rule is when researchers wish to use the same stimulus to assess repeatability (Hauber et al. 2014, Argument 1). In this case, within a subject one might wish to keep the stimulus constant. The generalizability issue is still present, however, and so multiple stimuli should still be used between subjects. A more controversial feature of such studies is the use of artificial stimuli to represent natural stimuli, which is a separate issue.

**Substitutes for Natural Stimuli**

An artificial stimulus is by its very nature an obstacle to generalization; so, all things being equal, a natural stimulus would of course be preferable. However, researchers might wish to modify natural stimuli or find or create substitutes, for several reasons. One, as previously mentioned, is for repeatability in studies where stimuli are destroyed in the course of the experiment. Another is when researchers wish to use experimental rather than statistical means of distinguishing the effects of different features of the stimulus (Hauber et al. 2014; Argument 2). More generally, appropriate natural stimuli are not always available. For instance, natural cuckoo eggs in some systems can be too uncommon to yield an effective sample size for egg rejection studies. In such cases, whatever the reason for their use, alternative stimuli (whether artificial or natural) can substitute as long as the subject cannot tell the difference—or, more formally, as long as behavioral responses to the two kinds of stimuli do not differ in ways that undermine the substitution.

**Generalizing from a Sample of Stimuli to the Whole Class**

The limits of artificial stimuli to represent natural stimuli are not always clear, but in general its objective is an experimental demonstration, rather than presumption or argument, that an artificial stimulus is perceived by the subject the way a researcher intends. Following are brief summaries of how the umwelt gamble can be appreciated and tested in experiments that apply artificial stimuli in different ways.
hair dye to simulate brightness of plumage in a seminal study, without controls for the substitution, although substantial subsequent research has thoroughly vindicated his assumptions in that paper (Hill 2002). Such success stories should not be taken as excuses for presumption, however—the effectiveness of a substitute stimulus is best defended by a direct empirical test.

Modifications and Exaggerations of Natural Stimuli

In some cases, researchers actually intend to present stimuli that the animals have never experienced in their lives. Certain research questions, for instance relating to the extent of behavioral flexibility or plasticity, can only be answered using stimuli that are modified or exaggerated from the range present in nature (Hauber et al. 2014; Argument 4). Here is where we meet Tinbergen and his supernormal stimulus, for example (Tinbergen 1951). Two mistakes to avoid when designing such experiments are (1) varying more than one feature of the stimulus at the same time without a control and (2) exaggerating from the natural range of stimuli to such a degree that the subject fails to consider the stimulus in the appropriate context.

In dozens of studies, two features of a stimulus are modified at the same time, and the researcher considers only one of these to be the cause of the response. However, when natural eggs with spots are painted white to test for the effect of spots on crypsis, one cannot be sure whether the difference in predation is due to the lack of spots or to the use of paint. Likewise, if adding spots to eggs leads to a negligible change in rejection rate, this could mean that spotting is not very important to rejection, or else that artificial spots do not elicit rejection as readily as natural variation in spotting does. If rejection of ‘mimetic’ conspecific natural eggs is compared with that of ‘non-mimetic’ eggs made of wax or plaster, or of painted eggs, the resulting difference in response might be due to the intended difference in egg appearance, or else to the artificial substance. For a final example, creating larger than normal spots with a Sharpie® permanent marker is a supernormal stimulus in one feature (spot size), but is also a novel stimulus with respect to reflectance due to the difference in pigment (e.g., xylene/cresol/PermaChrome® rather than protoporphyrin IX). Arguments to reviewers and article readers notwithstanding, nobody, including the researchers, can ever know which of the two simultaneously modified features is the salient one until they are tested separately. For instance, the researcher in the first example could present to predators a white natural egg without spots, or else a painted egg with spots, to separate paint from spots. In the last example, a sample of eggs with marker spots that are the same size as naturally occurring spots could be compared with naturally spotted eggs to test whether the response is a result of spot size or pigment.

A related danger is that an exaggeration or modification might be perceived by a subject as a wholly novel stimulus, in the sense of not being considered in the same context or as the same sort of trait as the natural stimulus it is supposed to be exaggerating or modifying. The relevance of an exaggerated stimulus can be tested by employing a set of stimuli that grade the exaggeration continuously; a continuous function among the responses from the subjects is expected. As with signals in tests of species recognition, the point at which a subject fails to respond predictably to a stimulus variant is the limit to legitimate supernormality or other stimulus modification.

One special case of researchers employing uninterpretable stimulus modification is in tests of the effect of ultraviolet reflectance (UV). The response elicited from a subject by a stimulus from which UV reflectance has been blocked entirely does not permit interpretation of the subject’s response to the UV component of the unmodified signal. The animal is responding to the eradication of an arbitrary portion of a stimulus. How this response relates to the desired one is unclear. Even if UV were largely redundant or unimportant in the identification of objects for a particular animal under normal circumstances, the experimental deletion of UV from an object’s reflectance might cause such a drastic change in its appearance that the animal’s identification ability could be disrupted or its response otherwise changed. The majority of tests of the importance of UV in behavioral research succumb to this pitfall.

Wholly Novel Stimuli

In some cases, artificial stimuli are not intended as a mimic, exaggeration, or graded modification of anything in nature. For instance, consider a study where the goal is to prompt a host to reject an egg, in order to determine the mode of rejection. One way to ensure rejection is to color the eggs extremely, in a manner the bird will be certain to consider foreign. The biological relevance of a wholly novel stimulus is the most difficult to defend, because of the higher stakes of the umwelt gamble and the difficulty of determining its outcome. With increasing unnaturalness or idiosyncrasy of a stimulus comes an increased likelihood that something about the artifice is going to
be perceived or interpreted by the subject in an unforeseen manner. Nancy Burley’s discovery during a study of female choice in zebra finches (Taeniopygia guttata) has become iconic in behavioral biology. By marking males with differently colored leg bands, she was inadvertently manipulating their attractiveness, as females preferred some colors over others; such preferences were later confirmed among wild birds as well (Burley 1988). Unexpected influences of color have also been found to enhance human performance in sports (Hill & Barton 2005). Tinbergen wondered why artificial green eggs were treated differently than black eggs in his experiments and postulated that the fact that leaves are green might have contributed (Tinbergen et al. 1962). To dust off another ethology term, we can describe this situation as latent valence, where ‘valence’ is a skew or bias in the response of an animal to a stimulus. Novel stimuli run the risk of tapping into unknown features of the infrastructure of perception and cognition that influence behavioral decisions. These features include those considered in contemporary behavioral sciences as learning predispositions, pre-existing or sensory biases, and behavioral reaction norms.

Consider the application of creating a ‘neutral abnormal’ stimulus; that is, one that is intended to elicit a response that is devoid of the specificity of any ‘normal’ stimulus. Such stimuli are expected to perform as extremes or nulls in some sense. Unfortunately, we do not know enough about an animal’s perceptual and cognitive apparatus to be sure what kind of stimulus is neutral, or extreme in an unbiased way, if any such thing exists. The biological perspective on behavior has taught us that classic psychology was wrong in treating all stimuli as equally learnable; equipotentiality appears to be rare if it exists at all (Ohman et al. 1976; Dall et al. 2005). An outlandish stimulus might elicit a response due to some latent valence of which we might be completely unaware. For instance, are highlighted or black painted eggs in a nest simply perceived as foreign and deserving of rejection, or do they interest the host or otherwise elicit peculiar responses? How different does an egg have to be from previously encountered variants before it is no longer considered an egg but simply a foreign object in the nest and perhaps treated in a way that is not entirely relevant to brood parasitism? This question holds even if, as has been suggested, egg rejection is evolutionarily derived from nest sanitation (Guignen & Sealy 2012). The question, stated a different way, is whether a highlighted or painted egg could be perceived by the animal as anything beyond what is intended by the researchers—anything but an egg that is maximally deserving of rejection. Researchers might assume that the answer to this question is ‘no’, whereas critics might not endure that assumption. Fortunately, there is a way to test it: A lineup of multiple artificial stimuli of wildly varying features will elicit similar responses if they are all neutral unfamiliar stimuli. So those glowing yellow eggs, those jet black eggs, those painted plaster eggs, and dice, can all be presented to subjects with the expectation that they are all equally prone to rejection. If highlighted eggs are rejected at different rates than black eggs, then color matters (beyond naturally existing variation). If, all other parameters being equal or controlled, the plaster eggs are rejected at different rates than the painted ones, then material matters. If the dice are rejected at different rates than the real and model eggs, then shape matters. But if they are all rejected at the same rate, this supports the assumption of neutrality of unfamiliar stimuli, within the range so far presented. This particular test has never been performed, although a few studies have begun to explore unfamiliar stimuli by describing patterns in the rejection of artificial eggs of a few colors in relation to their chromatic contrast (Cassey et al. 2008; Croston & Hauber 2014; Hauber et al. 2014; Moskát et al. 2014).

Conclusions

The use of artificial stimuli in behavioral research is widespread and valuable. However, the methods of existing studies (e.g., Fig. 1 and Appendix) are uneven and often elicit controversy because of the issues raised above. Whether the use of artificial stimuli is worthwhile or effective in a particular situation requires a cost-benefit analysis, and its use demands careful and controlled experimentation in order to produce interpretable results. Whether artificial stimuli succeed or fail boils down to whether the methodology successfully navigates the peculiarities of the study organism. The considerations above reduce to two general recommendations for research involving artificial stimuli: (1) reduce the umwelt gamble as much as possible, and (2) determine whether the gamble has succeeded through artifact detection tests.

As the goal in artifact detection is to rule out spurious results, statistics in such tests should be biased heavily toward the elimination of type II error (failing to detect an actual effect)—and instead to have high statistical sensitivity and low statistical specificity. Any practical possibility of an artificial result should call the respective method or stimulus into serious doubt.

The artificial manipulation of stimuli has a difficulty in common with many other areas of pure
and (especially) applied biology, such as genetics, medicine, biological control, and ecological restoration. The similarity among these circumstances is that we are toying with a system under the assumption that we understand it well enough to gamble that our strategy will work—that our creative manipulation is doing what we think it is doing. Commonly in these cases, the real world flouts our hopes and assumptions. My pilot study with wooden eggs failed, not because of the use of artificial or abnormal stimuli per se, but because my particular application of them fell foul of unforeseen aspects of my study organism, reducing any interpretation to triviality.

However, another take-home message of both Hauber et al. (2014) and this paper is that artificial stimuli can have distinct advantages. In another study population of weavers, this one having been introduced to the Dominican Republic, natural weaver eggs were not variable enough in color for me to present the same degree of variation to them in egg replacement experiments that was available in the ancestral population in Africa (Lahti 2005). To test for a change in egg recognition following introduction, I needed to present the introduced population with the same range of stimuli as I had presented to the ancestral population, so I painted the eggs. Colors did not match in the UV, of course, but an artifact detection test (in this case, painting eggs the same color as natural eggs) showed no effect of paint (Lahti 2006). Without recourse to artificial stimuli, I could not have tested for a historical change in egg recognition unless I (expensively and illegally) carted eggs across the Atlantic Ocean. So I am in agreement with the spirit of Hauber et al. (2014) that encourages peer reviewers to be careful and reasoned in their consideration of artificial stimuli in behavioral research. Whether this means more or fewer acceptances of such studies remains to be seen, however, and should depend at least as much on the diligence of researchers as of reviewers. What with zebra finches choosy about colored leg bands (Burley 1988), motmots fearing striped sticks as snakes (Smith 1975), and bees doubtful that a yellow spiderweb could be all that bad (Craig 1994), we have to appreciate that we are taking an umwelt gamble when we pick up that paintbrush or magic marker.

**Literature Cited**


Appendix

Fifty-four recent (2009–2014) brood parasitism studies employing artificial eggs or coloration


Guigueno, M. F. & Sealy, S. G. 2012: Increased investigation of manipulated clutches suggests egg recognition without rejection in a brown-headed cowbird (Molothrus ater) host, the yellow warbler (Setophaga petechia). Auk 129, 17—25.


