RESPONSE

Functional Variability, Matching, and Volition: A Reply to Davison (2007)

PsyCrit, December 8, 2007

In our recent publication, *Stochastic matching and the voluntary nature of choice* (Neuringer, Jensen & Piff, 2007), we sought to describe, using experimental methods, those characteristics of behavior that human participants would identify as "voluntary." We hypothesized that, even in a highly abstracted environment that provides little information, participants could consistently judge the appearance of volition. With this in mind, we showed the participants a number of different game-playing "actors" whose choices and reinforcers were controlled by simple algorithms. The participants were asked to rate the extent to which each actor's responses represent a voluntarily responding human player. Three conclusions emerged:

a) Participants can readily discriminate between certain types of response patterns (generated by different algorithms).

b) Participants show consistent trends in which behaviors they associate with "human-like" and "volitional."

c) The algorithmic actor most often identified as human/volitional chose among the available options in a stochastic fashion *and* in such a way that choices matched reinforcements.

Davison's PsyCrit critique of our work raises several cogent methodological points, but ultimately fails to address either our core hypothesis or our primary conclusions.

The first of Davison's (2007) two main criticisms refers to an ambiguity in our instructions to participants. He correctly points out that we routinely conflated "volitional" with "human-like," and asserts that our results are difficult to interpret because it is unclear which of these descriptors is the controlling stimulus. We used both descriptors intentionally because a common intuition, both among most philosophers and some of our participants, is that only humans have the capacity for free will (often equated to volition). Asking participants to rate "volition" in the absence of "human-ness" is a contradiction in terms for at least some people, and one that we did not wish to confront. At the same time, human behavior is not *necessarily* volitional. People sometimes act compulsively, reflexively, mindlessly, or insanely, and therefore we did not think that appending "human" to "voluntary" would bias our results. Teasing "human" and "voluntary" apart can be an objective for future research, although doing so promises to be difficult.

Davison's second criticism is more fundamental, and implies that our findings themselves are evidence of a flawed methodology. To discuss this criticism, we must briefly describe the basic procedure and results in somewhat more detail. Our participants watched actors, represented on a computer screen, repeatedly choose among three gambling alternatives, each of which was intermittently reinforced, often with different probabilities. Technically, the actors were reinforced under concurrent dependent random-ratio schedules of reinforcement (see Jensen & Neuringer, submitted; Lau & Glimcher, 2005). Different actors responded in different ways, all being programmed by computer-based algorithms based on Baum's Generalized Matching Law (Baum, 1974). The different actors represented different values of the *sensitivity* [*s*] exponent of that function. Some algorithms tended to generate responses randomly, with relatively equal probabilities across the three alternatives no matter the distributions of received rein-

forcers (undermatchers with $s < 1.0$). Other algorithms tended to generate extreme preference for the most frequently reinforced option, despite the fact that a higher frequency of reinforcement could be obtained by distributing choices more widely (overmatchers, with $s > 1.0$). A third algorithm, the "matcher" ($s = 1.0$), tended to distribute responses in stochastic fashion so as to match obtained proportions of reinforcers. The main result was that human participants judged the stochastic matchers ($s = 1.0$) to approximate most closely the performance of a voluntarily choosing human.

Davison's criticism is that real humans rarely match (we use "match" to refer to the $s = 1.0$ case) whereas matching was perceived as best representing human volition in our experiments. Davison cites an excellent review paper by Kollins, Newland and Critchfield (1997) in support. When faced with concurrent-reinforcement contingencies similar (although not identical) to those employed in our experiments, humans only occasionally match. However, we do not claim that humans *always* match, or that they are compelled to match, or that voluntary actors *must* match. We readily accept all of the examples provided by Davison (2007) and by Kollins et al. (1997) of non-matching behaviors in humans (and other organisms). But we join Kollins et al. in hypothesizing that the different types of matching performances seen in experiments with humans, as well as non-human animals, may be related to different "establishing operations," or contexts. In situations where matching is punished or where non-matching behavior is necessary for reinforcement, humans may *not* match. In other cases, humans may fail to match because the context is too complex or because events are difficult to discern. In the present experimental situation, however, matching *was* functional (adaptive), and we submit that functionality is *one* characteristic of volition.

In a related criticism, Davison argues that Baum's generalized matching function does not adequately describe choice behaviors and he offered an alternative function. We do not know how human participants in our experiments would have responded to the alternatives, but the Baum formulation sufficed as a model to test our hypothesis that functional variations in allocation of choices contribute importantly to high judgments of volition.

Of a number of control conditions described in our paper, two are most important for this discussion. First, we found that matching *alone* was not sufficient to generate high ratings from our participants. In one experiment, one actor conformed to the matching law, but did so in such a way that its responses were structured and therefore predictable. It was rated as considerably less volitional than the stochastic matcher despite achieving the same rate of reinforcement. Second, judgments of volition were not based solely on attainment of most frequent reinforcement, as algorithms that exceeded the stochastic matcher in efficiency were nevertheless rated as less human. Again, stochastic (and therefore sometimes unpredictable) allocation of responses was required.

Stochasticity was ignored by Davison, but it is critical to our argument. To be more precise, our most consistent result, across all of our experiments, was that *stochastic* matching was perceived to best represent human-like volition. In one experiment, employing a Turing-type test, participants were told that some of the actors represented *real* human voluntary responders. In this experiment, the stochastic matcher was identified as a voluntarily responding human more often than any other algorithm. Stochastic matching is characterized by a response strategy in which choices are generated in a probabilistic fashion, and, at the same time, overall frequencies of response match frequencies of obtained reinforcers. We hypothesize that stochastic matching was perceived as best representing human volition because of two behavioral characteristics: (a) functionality of choice distributions and (b) levels or degrees of variability (or predictability) that functionally change, depending upon the reinforcing context. Regarding the first of these, an observer must be able to identify the behavior as satisfying some goal-oriented objective (or being directed at obtaining a potentially reinforcing outcome) for the behavior to be judged voluntary. Regarding the second, in some situations, highly repetitive and predictable choices are ap-

propriate whereas in others, unpredictable responding is best. The ability to go from predictable to unpredictable, and do so in a functional way, is a necessary characteristic of voluntary behavior and indicates, to an observer, the ”freely willed” nature of the response.

Our central finding, that the interaction of unpredictability and functionality leads to volitional judgments, should come as no surprise to those familiar with operant variability research. As argued by Neuringer (2002), unpredictable responding is the most functional approach an organism can take in some scenarios. From the “mixed strategies” central to modern game theory (Glimcher, 2003), to the explicit reinforcement of random-like behavior (Neuringer, 2002), to the benefits of unpredictability in predator evasion (Maynard Smith, 1982), goal-oriented behavior in the real world must be variable when the situation demands. We propose that volition can best be understood as the interplay between an organism’s level of variability and the functionality of its behavior.

One consequence of this characterization is that rather than being as all-or-nothing, volition can be considered as a continuum. Some behaviors can be described as more voluntary than others (a mentally healthy adult’s behavior vs. someone with a pathological compulsion, for example), and humans can be contrasted in their level of volition to that displayed by animals or artificial life (such as computer programs). Our methodology of asking for ”levels of approximation to voluntary human choices” in some of our experiments extends from this idea that volition can be continuous rather than nominal.

One additional point raised by Davison is that we studied only perceptions, not volition itself. Our focus on *perceptions by external observers* of another’s voluntary *behaviors* is quite uncommon in the history of the study of volition and free will. We indicate, and Davison repeats, that agreement as to the perceived characteristics of volition is not itself evidence for the reality of volition. Agreement as to what Santa Claus looks like does not itself provide evidence for the reality of Santa. Perhaps we misstated our case, however, since such agreement *can* provide evidence. Santa Claus *is* real, in the sense that he or she walks the streets on Xmas, often appears in homes, and is instantly recognized when dressed in a certain red costume and wearing a white beard. What is *not* real is the explanation, or interpretation, of that reality. Santa is not an other-worldly being who descends in a sleigh from the heavens. Similarly, we take agreement concerning voluntary behavior to indicate something about the reality of behavior-environment relationships, whether or not the explanations and theories about volition have been in error.

In conclusion, we feel that our results may contribute importantly to the free will debate, despite Davison’s concerns. Our findings show that people attribute human-like volition to certain behaviors-in-contexts and this in turn creates a new springboard from which the debate can continue.

Greg Jensen
Allen Neuringer
Reed College
allen.neuringer@reed.edu

REFERENCES

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, *22*, 231-242.

Davison, M. (2007, October) Is it human? Judgments of choice responding. PsyCrit. Retrieved December 2, 2007, from http://psycrit.com/wikiup/b/b8/DavisonNeuringer2007.pdf .

Glimcher, P. W. (2003). *Decisions, uncertainty, and the brain.* Cambridge, MA: MIT Press.

Jensen, G. & Neuringer, A. (submitted). Choice as a function of reinforcer "hold": From probability learning to concurrent reinforcement.

Kollins, S. H., Newland, M. C., & Critchfield, T. S. (1997). Human sensitivity to reinforcement in operant choice: How much do consequences matter? *Psychonomic Bulletin & Review, 4*, 208-220.

Lau, B. & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 84*, 555-579.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin & Review, 9*, 672-705.

Neuringer, A., Jensen, G., & Piff, P. (2007). Stochastic matching and the voluntary nature of choice. *Journal of the Experimental Analysis of Behavior, 88*, 1-28.