Preface

SQAB 2006: "It's the Non-Arbitrary Metrics, Stupid!"

SQAB began as a symposium held every year at Harvard University, and the proceedings of the early meetings were published as a series of volumes edited by Michael Commons, Richard J. Herrnstein, John A. Nevin, Howard Rachlin, and others. Initially each conference was organized around a specific theme, such as reinforcement delay (Commons et al., 1987), reinforcement schedules and discrimination (Commons et al., 1983), matching versus maximizing (Commons et al., 1982), and signal detection (Commons et al., 1991). Since 1993, SQAB has held its annual meeting in conjunction with the Association for Behavior Analysis. Due to efforts by SQAB's program chair at the time, William Palya, and continued by his successor Armando Machado, individual programs became more eclectic, frequently including presentations from researchers in diverse fields such as economics, decision making, neural networks, and behavioral ecology.

The increasing breadth of topics shows the applicability of SQAB's methodology – i.e., the quantitative analysis of behavior – to substantive problems in a broad range of topics in psychology and related disciplines. Perhaps SQAB's most distinctive features are its analysis of the behavior of individual organisms, eclectic use of a wide range of quantitative methods, and an emphasis on developing mathematical models. In effect, SQAB has evolved by integrating various perspectives from researchers interested in the behavior of organisms, including Skinner's radical behaviorism, comparative and cognitive psychology, economics and biology, with the full range of mathematical and quantitative methods available to modern science, including formal-analytic, statistical (inferential and descriptive) and computational approaches.

But what about the rest of psychology? Even a cursory examination of trends over the last few decades shows that quantitative approaches have become more sophisticated. Many journals now require authors to report confidence intervals and statistical power along with results of traditional significance tests, meta-analysis is regarded as the standard method for integrating primary research literatures (Rosenthal and DiMatteo, 2001), and structural equation modelling and related approaches are increasingly used by researchers in traditionally 'softer' areas like social and personality psychology (e.g., Olsen and Kenny, 2006), facilitated by the widespread availability of sophisticated software packages like SPSS and Statistica. Could it be that the rest of psychology is becoming, for want of a better phrase, more SQAB-like?

In fact, there are some critical differences between the quantitative methodologies used by mainstream psychology and those associated with SQAB. A recent article in *American Psychologist*, the flagship journal of the American Psychological Association, raises several troubling issues for mainstream psychology that are relevant here. Blanton and Jaccard (2006) criticized the use of 'arbitrary metrics' in psychology. They defined an arbitrary metric as a measurement in which "it is not known where a given score locates an individual on the underlying psychological dimension or how a one-unit change on the observed score reflects the magnitude of change on the underlying dimension" (p. 28). They provided several examples of the use of arbitrary metrics, including the Implicit Association Test (Greenwald et al., 1998), in which response latencies on a computerized task are used to infer the degree of an individual's hidden prejudices and stereotypes, and the notion of 'clinical significance', in which an individual's change on some psychologically-relevant dimension during treatment (e.g., depression) is said to be clinically important or meaningful if their post-treatment score is within a 'normal' range defined for a non-clinical population. According to Blanton and Jaccard (2006), arbitrary metrics are not problematic when their purpose is to test theoretical predictions. For example, the distance that a male research participant chooses to sit from a female confederate in an experiment may be used as a proxy for 'interpersonal attraction', and used to test a prediction that perceived similarity should result in increased attraction. However, it would be a mistake to infer that meaningful statements can be made about the absolute degree of attraction, as an underlying psychological construct, experienced by a male participant in the experiment. But such inferences are commonly made; Blanton and Jaccard noted that approximately 73% of visitors who complete the IAT on a popular webpage are told that they have 'slight', 'moderate', or 'strong' racial prejudices, based on a standardized effect size measure,[1] even though there is

---

[1] The effect size measure used in the web-based IAT is based on Cohen's *d*, and is calculated for each individual as the difference between the average response latencies for the two conditions divided by the individual's standard deviation. Thus, the effect size metric is arbitrary because even though it is calculated from reaction times, it is rescaled for each individual depending on their standard deviation. A 'large' effect can be obtained either by having a greater difference in mean response latency or by more consistent responding.

no evidence linking this measure to differences in real-world behavior.

Of course, the examples discussed by Blanton and Jaccard (2006) represent just the tip of the iceberg as far as the use of arbitrary metrics in psychology is concerned. But it is important to note that why they are so problematic is related to the distinctive features of SQAB discussed above. The first problem with arbitrary metrics might be termed *faith in measurement meaning*: the assumption that the scale of the dependent variable corresponds in a meaningful way to the scale of the psychological construct. The requirements for meaningfulness have been examined in detail, but are rarely met by psychological measurement (Michell, 1990; Narens, 2001). To its credit, SQAB does not take meaningfulness as an article of faith. Its dependent variables are observable aspects of behavior and environmental stimuli that are measurable in quantitative terms—count-based measures derived from absolute or relative frequencies of events (e.g., a pigeon's key peck or rat's lever press), time-based measures (e.g., time allocated to different activities or response latencies), and other physical measures (e.g., amount of food or water consumed). The second problem noted by Blanton and Jaccard (2006) is the use of norming against a reference group, particularly in the context of "clinical significance". Because SQAB has always emphasized the behavior of individual organisms, the pitfalls associated with expressing an individual's score relative to a population are avoided.

The appearance of Blanton and Jaccard (2006) article in such a widely-read journal is good news, and shows that the issues of meaningful measurement, essential to science, are being taken seriously in mainstream psychology. But SQAB has always understood this. By combining use of non-arbitrary metrics with innovative quantitative methods, the research tradition represented by SQAB has been able to make a substantial contribution to psychology as well as related disciplines. The articles in this special issue, which describe research presented at the 29th annual meeting of SQAB in Atlanta, Georgia, May 26–27, 2006, enhance this contribution. New for this year's special issue is a section of Short Communications, which are based on presentations at the poster session.

Researchers in the SQAB tradition have had considerable success studying the variables that influence choice. They have developed detailed procedures such as concurrent and concurrent-chains schedules and temporal discounting procedures, and generated a variety of mathematical models that aid our understanding of choice behavior of both humans and non-humans. Some of the best-known results in this area are that response rate is a hyperbolic function of reinforcer rate (Herrnstein, 1970), and that response allocation in concurrent schedules is well-described as a power function of reinforcer allocation with an exponent less than 1 ('undermatching'; Baum, 1979). McDowell and Caron describe the results of simulations that show that a new model, based on a genetic algorithm and incorporating principles of selection, mutation, and reproduction, is able to explain both the hyperbolic response-rate function and undermatching as emergent phenomena. Their conclusions are based on a rigorous quantitative analysis and suggest that their new model may ultimately provide a complete mathematical dynamics of behavior.

In the last few decades, some behavior analysts have extended their studies of choice to the area of human judgment and decision-making, joining investigators in other fields such as economics, cognitive psychology, behavioral ecology, and social psychology. One productive technique of studying decision making in the laboratory has been to expose participants to economic games in which their choices produce different payoffs. The article by Fantino, Gaitan, Kennelly, and Stolarz-Fantino uses this approach and continues a highly successful record of research in this area by Fantino and his colleagues. The central issue of this paper is to understand how the type of payoff influences choice. Behavioral economists have argued that real money, rather than hypothetical rewards, should be used in these economic games if the results are to have generality. Yet, experiments that provide real money can become expensive to carry out. In two experiments involving different economic games, Fantino et al. find ample differences in choices made when real versus hypothetical rewards are used, thus supporting the concerns expressed by behavioral economists. Luckily for behavioral science, they also find that time off from a tedious task can serve as a powerful reinforcer just as effective as money. This intriguing finding offers the promise that behavioral research in this area need not be prohibitively expensive to carry out.

The study of temporal discounting may be the area of human decision-making research in which the issue of real versus hypothetical reward is most important. Temporal discounting procedures estimate the reduction in the subject's subjective value of a reward as a function of delay. The discounting of delayed rewards is typically measured by providing the subject with a choice between a relatively immediate, smaller reward and a more delayed, larger reward. When humans are used as subjects, the delay discounting assessment task presents the two options on paper or on a computer screen. One option is a certain amount of hypothetical money now, versus a larger amount of hypothetical money (perhaps thousands of dollars) provided at varying periods up to years into the future. This task is designed to determine the characteristics of human decision-making in realistic situations when one's decision may produce a payoff many years later, such as the decision to obtain a university degree or invest in the stock market. Hypothetical rewards are necessarily used in this type of research because real money simply could not be given with the delay periods utilized.

Two articles published as Short Communications in this special issue compare procedures used to assess temporal discounting in humans. Robles and Vargas continue their work in this area by asking whether presenting the choice values in ascending, descending, or random order would produce the same indifference points representing the same amount of delay discounting. In a within-subject design, they observed that the ascending and descending procedures produced equivalent delay discounting gradients, but providing the choice values in random order greatly increased the steepness of the discounting gradient. All three procedures produced gradients that were more accurately described by a hyperbolic model than by an exponential model.

In the second article, Kowal, Yi, Erisman, and Bickel compare two algorithms used in computerized temporal discounting procedures with humans. The central issue is whether the Decreasing Adjustment algorithm would produce the same amount of temporal discounting as the Double-Limit algorithm, or even if they measure the same behavioral process. These two computer algorithms differ in the way in which they adjust the value of the immediate outcome (hypothetical money, gains and losses) from trial to trial in each choice option provided to the subject, with the goal of approaching the point at which both outcomes are subjectively equivalent. The Decreasing-Adjustment algorithm produced lower indifference points and higher rates of discounting than the Double-Limit algorithm. Nevertheless, strong correlations observed with the two procedures support the claim that the two procedures do, indeed, measure the same behavioral process.

The Short Communication by Maguire, Hughes and Pitts is also relevant for our understanding of choice behavior. They studied responding of pigeons in a 'rapid acquisition' concurrent schedules procedure in which reinforcer magnitude changed unpredictably across sessions. Previous research has shown that response allocation can track changes in reinforcer rate (Schofield and Davison, 1997) and immediacy (Grace et al., 2003), effectively providing a learning curve within each individual session. Maguire et al.'s experiment demonstrates that pigeons' response allocation can track changes in reinforcer magnitude as well. This result is potentially important because it shows the rapid acquisition procedure might provide a useful baseline for studying the effects of drugs or other interventions on choice between alternatives that differ in terms of reinforcer magnitude.

The study of timing has always been one of the most important areas in the quantitative analysis of behavior. Several articles in the special issue ask fundamental questions about the nature of timing. Odum and Ward review the most popular and influential account of timing – scalar expectancy theory (SET; Gibbon et al., 1984) – and the assumption that the biological substrates of SET's pacemaker–accumulator system involve dopamine (Meck, 1996). Noting that results of previous studies which have tested whether psychometric functions are shifted to the left through administration of dopamine agonists (e.g., *d*-amphetamine) are mixed, Odum and Ward note several important procedural differences that might account for the mixed results. In particular, studies using temporal discrimination procedures have differed according to whether the choice responses have fixed location ('position task') or are associated with different colors that change location ('color task'). They go on to present several experiments which provide a within-subject comparison of the effects of *d*-amphetamine on the psychometric function in the two tasks. For both color and position tasks, they found that *d*-amphetamine produced a disruption of temporal discrimination, as evidenced by a flattening of the psychometric function. However, there was no evidence of a consistent leftward shift, as predicted by the SET's dopamine hypothesis.

Like the article by Odum and Ward, the Short Communication by Guilhardi, MacInnis, Church, and Machado measures shifts in the psychophysical functions that relate response rate with time since trial onset. One goal was to test predictions from two theories of timing, SET and Learning to Time (LeT: Machado, 1997). A second goal was to compare the results from their rats to previously published results using pigeons (Machado and Guilhardi, 2000). Sixty-second trials were divided into four segments with VI schedules that differed in local reinforcement rate, while holding overall reinforcement rate in the trial constant. Responses on the left lever were reinforced during the first half of the trial (first two segments), and responses on the right were reinforced during the second-half. Manipulation of local reinforcement rate around the middle of trials was sufficient to produce shifts in the psychophysical function. Subjects (rats and pigeons) shifted earlier in the trial from the early response (the left lever or key) to the late response (right) if the late response was reinforced at a higher rate near the point of transition. The authors conclude that the shifts in psychophysical functions are more compatible with explanations based on LeT than on SET.

Comparing predictions of SET and LeT is also a focus of the Short Communication by Linlin Yi. She analyzes responding of rats during no-food trials from a multiple peak procedure. Predictions of a recent dynamic model for timing, Packet theory (Kirkpatrick, 2002), are also considered. In terms of goodness of fit, all three models performed remarkably well, accounting for approximately 99% of the variance in the group-mean data. This result indicates that selecting the best model will require more than a simple comparison of explained variance, and Yi presents a sophisticated analysis of not only the residuals from each model, but an attempt to measure model complexity through a cross-validation analysis (cf. Myung, 2000). This shows that the model which provided the best overall account in terms of goodness of fit, Packet theory, suffered from over-fitting because its predictions did not generalize as well as the other models in a cross-validation sample. Although the results do not allow us to choose conclusively which model is best, Yi's analyses demonstrate the degree of rigor that is possible (indeed, necessary) to test contemporary models for timing.

The Short Communication by MacInnis also investigates characteristics of animal timing. In this case, the key question is whether rats time "empty" and "filled" intervals of equal duration differently. Filled intervals are tracts of time containing a continuous presentation of a stimulus, whereas empty intervals contain only a brief stimulus to signal the onset and offset of the interval. William James, in his 1890 psychology text, may have been the first to point out how we perceive filled intervals to last longer than empty intervals. Research with humans has repeatedly confirmed this finding, but research with rats has not been as common. MacInnis presented 30-s filled or empty intervals at the last part of a fixed-interval (FI) 120-s schedule that required head entries into the feeder opening. White noise was used as the brief or continuous stimulus in these intervals. The rats showed better temporal discrimination on the filled intervals than on the empty intervals, as predicted by studies with humans. The rats waited longer to start responding during the filled intervals, indicating that they were timing more precisely. These results encouraged a comparison between two competing models of animal timing. SET explains these results by adjusting the pacemaker rate in the internal clock, and Packet Theory is able to account for the

data without changing any of the assumptions about the internal clock for the two interval types.

Questions related to timing are also addressed in the Short Communication by Caetano, Guilhardi and Church. They present a secondary analysis of data from an experiment by Guilhardi and Church (2005) in which rats were trained on a multiple fixed-interval procedure (with intervals of 30, 60, and 120 s). Different groups of rats were trained either with all three intervals presented in each session ('simultaneous condition'), or with the different intervals presented across blocks of sessions ('blocked condition'). Caetano et al. found that although the two groups responded similarly during baseline training, their behavior differed in a subsequent transfer test. These results are important because they challenge assumptions of current models of timing, including SET, LeT, and Packet theory, that responding under the simultaneous and blocked conditions should result in equivalent learning.

Research in behavioral pharmacology commonly characterizes drug effects in terms of their influence on overall response rate. Yet when behavior analysts discovered substantial structure in behavior at the molecular level of interresponse intervals (IRTs) and response bouts, behavioral pharmacologists asked whether restricting analysis to overall response rate would overlook the effects of drugs acting at this molecular level. Two articles published as short communications in this special issue examine how drugs act at the molecular level of IRTs or the dynamics of switching between response alternatives.

The key question asked by Bennett, Hughes, and Pitts in their paper in this special issue was whether microstructural analyses would reveal order in the effects of methamphetamine administration to pigeons that might not be evident in overall response rate. They answer in the affirmative by demonstrating that methamphetamine does not have a blanket effect on all IRT classes. They suggest that these different classes of IRTs may not be functionally equivalent under the effects of methamphetamine administration.

In a separate paper, Aparicio points out that if dopamine actually does mediate the reinforcing effects of stimuli, as widely assumed in the neurosciences, then dopamine antagonists should decrease the reinforcing value of food reinforcers. So, how can one measure the reinforcing value of food reinforcers? The generalized matching law describes choice between concurrently available reinforcement schedules. When these schedules differ in the values of the reinforcers, the parameter reflecting sensitivity to the reinforcer ratio (s) is believed to reflect the relative values of reinforcement in the two alternatives. Therefore, rats given haloperidol, a dopamine antagonist, and exposed to varying concurrent reinforcement schedules would be predicted to show systematic changes either in the sensitivity parameter, s, or the bias parameter, b. Aparicio assessed the effects of haloperidol on dynamic choices in schedules in which the frequency and magnitude of reinforcement changed simultaneously within and between sessions. Nevertheless, haloperidol did not have the expected effects. The match between predicted and obtained response ratios was not affected by the drug (at any of the four levels administered), which indicates that preference did not change. Haloperidol did not appear to decrease the "pleasure"

derived from the food reinforcers. Also, sensitivity (s) of the behavior ratio to changes in the reinforcer ratio reached levels that are consistent with those found in similar situations without the administration of haloperidol, and the parameter value was approximately the same in the control and vehicle conditions. Aparicio concludes that haloperidol does not decrease the motivation for food reinforcers nor suppress appetite.

Operant resurgence refers to an increase in the rate of behavior that was previously reinforced but subsequently extinguished, when an alternative response is extinguished. Procedures for studying resurgence generally include three (or more) successive conditions. The first condition reinforces a particular target response (or sequence); the second condition extinguishes that previously reinforced response and may reinforce a different one; and the third condition extinguishes the second reinforced response (and may reinforce a third one) while providing the opportunity for the first response to resurge without being reinforced (Bachá-Méndez et al., 2007). Resurgence is a topic of obvious relevance not only for our understanding of behavior in general, but also for those interested in developing effective interventions for problem behavior. Thus, it may seem surprising that there have been relatively few systematic investigations of resurgence. The article by Doughty, da Silva and Lattal goes some way towards addressing this gap. They present results from a series of experiments that attempts to isolate the critical determiners of operant resurgence relating to how the target response is eliminated. Their results show that resurgence of keypecking in pigeons is minimized when target responding is eliminated through reinforcement of an alternative keypeck response. This result has some noteworthy implications for applied behavior analysis, because it suggests that a response which is topographically different from the one which is currently being extinguished is most likely to show resurgence.

Many, if not most, of researchers in the SQAB tradition would endorse the perspective of 'general process learning theory', in which a representative organism such as a pigeon or rat is studied under well-controlled laboratory conditions to reveal principles of behavior which are expected to have widespread generality. Of course, 'expected' is the operative word and questions of whether results such as the matching law generalize across species or outside the laboratory context are obviously important. Two Short Communications are relevant to this question in different ways. James Mazur describes results of an experiment which suggests that choice behavior of rats and pigeons may be differentially affected by stimuli which signal probabilistic rewards. In his experiment, rats chose between a standard alternative that delivered food after a 5-s delay with a specified probability (e.g., 50%), and an adjusting alternative that provided food on 100% of the trials but after a delay that was titrated across trials to estimate an indifference point. The critical comparison was whether the different outcomes on the standard alternative (i.e., food or no food) were signalled by distinctive stimuli. Mazur found that rats' choices for the standard alternative were similar regardless of whether the distinctive stimuli were presented. This finding provides a strong contrast with results from related experiments with pigeons, which have found that preference for the probabilistic alternative increases if differ-

ent stimuli are associated with food and no-food outcomes (e.g., Mazur, 1989). Mazur suggests that rats' choices might be primarily determined by the delay between the response and reinforcer, whereas pigeons' choices are controlled by the delay between the stimulus and reinforcer. Whatever the ultimate explanation for this result might be, Mazur has identified an important species difference that needs to be further investigated, especially considering that rats and pigeons are the most popular non-human experimental subjects in behavior analysis.

Reptiles are far less commonly studied in the behavioral laboratory than mammals or avians, and so Davis and Burghardt's Short Communication is a welcome contribution. They describe research in which a turtle species native to Florida, *Pseudemys nelsoni*, was trained to perform a discrimination task. In the final version of the procedure, turtles had to swim to the opposite end of an aquarium, climb out of the water onto bricks, and knock over one of two clear plastic bottles that contained food. The turtles were trained to perform the task successfully, and demonstrated long-term retention when tested over 7 months later. Davis and Burghardt's results show that *Pseudemys nelsoni* can be used effectively in the behavioral laboratory, and future research should explore the abilities of this species further.

The Short Communication by Burgos and Murillo-Rodriguez continues the impressive series of computer simulations that demonstrate that the 'selection neural network' can reproduce important empirical phenomena. This neural-network architecture was first proposed by Donahoe et al. (1993), and it has successfully reproduced a wide variety of behavioral phenomena—especially those related to classical conditioning. In this article, the first of these demonstrations is that of context specificity, or the context shift effect: the decrease of conditioned responding when the conditioned stimulus is tested in a context different from the one in which it had been paired with the unconditioned stimulus. The other effect is renewal, the recovery of the conditioned response in the training context after extinction in another context. Their report ends with a general discussion of the virtues and vices of the strategy used to build the model and some caveats of the simulations.

Finally, there are two articles, by Gordon Foxall and Charles Shimp, which might seem atypical at first because they neither present any empirical data nor a new quantitative model. But these papers contribute to our research enterprise at a conceptual or meta-theoretical level, by organizing how we think about important phenomena and perhaps more basically, by just encouraging us to ask questions. Foxall describes a new framework for the analysis of intentionality that he has developed for behavior analysis. This may come as a bit of a surprise—what's a good behaviorist like Gordon Foxall doing with theoretical concepts like intentionality that reek of introspection and folk psychology? Not everyone may be convinced, but Foxall argues cogently that intentionality is a necessary explanatory ingredient for a behavior-analytic account of complex human phenomena. Foxall's empirical research program has already achieved impressive results in showing that phenomena such as consumer brand choice can be understood in terms of behavioral principles (e.g., Foxall et al., 2004); it remains to be seen whether his new framework will enable an even more powerful analysis of consumer behavior.

Charles Shimp's article asks a simple but fundamental question: what are the implications of turning the analytical, methodological, and explanatory tools of SQAB onto itself? In other words, what can we learn, if anything, from attempting to understand the behavior of SQAB researchers at a level consistent with their explanations of the behavior of other organisms? The crux of his argument is presented as a *gedankenexperiment*. Specifically, he claims that because radical behaviorism requires that the same principles must apply to the behavior of scientists as well as their laboratory subjects, that SQAB must provide an account of itself. Shimp's analysis is truly radical and not all will agree, but even if the 'reflexive behavior analysis' that he calls for turns out to be impossible, his paper makes a valuable contribution by encouraging researchers to reflect on their own activity as scientists.

The articles in this special issue underscore the major points made earlier. As a research tradition, SQAB is able to make cumulative progress because real behavior is measured in non-arbitrary metrics, aided by powerful quantitative techniques for model comparison, theory development and testing. The range and diversity of topics covered provides ample illustration of SQAB's eclecticism and relevance for different areas in psychology and related fields. And finally, we are encouraged by the young and emerging investigators who have contributed many of the Short Communications published here. They will develop into tomorrow's established researchers, and the high quality of their work is a positive sign for SQAB's future.

References

Bachá-Méndez, G., Reid, A.K., Mendoza-Soylovna, A., 2007. Resurgence of integrated behavioral units. J. Exp. Anal. Behav. 87, 5–24.

Baum, W.M., 1979. Matching, undermatching, and overmatching in studies of choice. J. Exp. Anal. Behav. 32, 269–281.

Blanton, H., Jaccard, J., 2006. Arbitrary metrics in psychology. Am. Psychol. 61, 27–41.

Commons, M.L., Herrnstein, R.J., Rachlin, H. (Eds.), 1982. Matching and Maximizing Accounts: Quantitative Analyses of Behavior, vol. 2. Ballinger, Cambridge, MA.

Commons, M.L., Herrnstein, R.J., Wagner, A.R. (Eds.), 1983. Discrimination Processes: Quantitative Analyses of Behaviour, vol. 4. Ballinger, Cambridge, MA.

Commons, M.L., Mazur, J.E., Nevin, J.A., Rachlin, H. (Eds.), 1987. The Effect of Delay and Intervening Events on Reinforcement Value: Quantitative Analyses of Behavior, vol. 5. Lawrence Erlbaum, Hillsdale, NJ.

Commons, M.L., Nevin, J.A., Davison, M. (Eds.), 1991. Signal Detection: Mechanisms, Models, and Applications: Quantitative Analyses of Behavior, vol. 10. Lawrence Erlbaum, Hillsdale, NJ.

Donahoe, J.W., Burgos, J.E., Palmer, D.C., 1993. A selectionist approach to reinforcement. J. Exp. Anal. Behav. 60, 17–40.

Foxall, G.R., Oliveira-Castro, J.M., Schrezenmaier, T.C., 2004. The behavioral economics of consumer brand choice: Patterns of reinforcement and utility maximization. Behav. Process. 65, 235–260.

Gibbon, J., Church, R.M., Meck, W.H., 1984. Scalar timing in memory. Ann. N. Y. Acad. Sci. 423, 52–77.

Grace, R.C., Bragason, O., McLean, A.P., 2003. Rapid acquisition of preference in concurrent chains. J. Exp. Anal. Behav. 80, 235–252.

Greenwald, A.G., McGhee, D.E., Schwartz, J.L.K., 1998. Measuring individual differences in implicit cognition: the Implicit Association Test. J. Pers. Soc. Psyc. 74, 1464–1480.

Guilhardi, P., Church, R.M., 2005. Dynamics of temporal discrimination. Learn. Behav. 33, 399–416.

Herrnstein, R.J., 1970. On the law of effect. J. Exp. Anal. Behav. 13, 243–266.

Kirkpatrick, K., 2002. Packet theory of conditioning and timing. Behav. Process. 57, 89–106.

Machado, A., 1997. Learning the temporal dynamics of behavior. Psychol. Rev. 104, 241–265.

Machado, A., Guilhardi, P., 2000. Shifts in the psychometric function and their implications for models of timing. J. Exp. Anal. Behav. 74, 25–54.

Mazur, J.E., 1989. Theories of probabilistic reinforcement. J. Exp. Anal. Behav. 51, 87–99.

Meck, W.H., 1996. Neuropharmacology of timing and time perception. Brain Res. Cog. Brain Res. 3, 227–242.

Michell, J., 1990. An Introduction to the Logic of Psychological Measurement. Lawrence Erlbaum, Hillsdale, NJ.

Myung, I.J., 2000. The importance of complexity in model selection. J. Math. Psychol. 44, 190–204.

Narens, L., 2001. Theories of Meaningfulness. Lawrence Erlbaum Associates, Mahwah, NJ.

Olsen, J.A., Kenny, D.A., 2006. Structural equation modelling with interchangeable dyads. Psych. Method 11, 127–141.

Rosenthal, R., DiMatteo, M.R., 2001. Meta-analysis: recent developments in quantitative methods for literature reviews. Ann. Rev. Psych. 52, 59–82.

Schofield, G., Davison, M., 1997. Nonstable concurrent choice in pigeons. J. Exp. Anal. Behav. 68, 219–232.

Randolph C. Grace [*]
University of Canterbury, Department of Psychology,
Private Bag 4800, Christchurch, New Zealand

Alliston K. Reid
Wofford College, Department of Psychology,
429 N. Church St., Spartanburg, SC, USA

[*] Corresponding author. Tel.: +64 364 2987 7996,
+864 597 4642.
E-mail addresses: randolph.grace@canterbury.ac.nz
(R.C. Grace),
Alliston.Reid@wofford.edu
(A.K. Reid)