Running Head: REASONING VERSUS ASSOCIATION

Reasoning Versus Association in Animal Cognition: Current Controversies and Possible Ways Forward

Tom Beckers: KU Leuven and University of Amsterdam

Jan De Houwer: Ghent University

Dominic Michael Dwyer: Cardiff University

5th February 2016 – Accepted for publication in the Journal of Comparative Psychology

Editorial correspondence to: Tom Beckers
Department of Psychology
KU Leuven
Tiensestraat 102, box 3712
3000 Leuven
Belgium
Tel +32 16 32 61 34 / Fax +32 16 32 60 99
E-mail tom.beckers@kuleuven.be

Abstract

The study of animal cognition is rife with controversy, and among the most long-standing and most intensely debated controversies in the field is the question to what extent the behaviour of non-human animals can be fully understood on the basis of purely associative principles, or whether some behaviours exhibited by animals necessitate the assumption of inferential capacities in animals that defy an associative explanation. Remarkably, the continuing debate on the topic seems to be spawning little genuine progress in terms of substantial accumulation of new, generally accepted, insights. As an introduction to a special section of the *Journal of Comparative Psychology* on the topic, the present paper outlines a number of reasons for the stalemate and suggests ways to re-fertilise the debate. In particular, we claim that progress will not come from the adoption of general principles like Morgan's Canon or the primacy of prediction over postdiction. Instead, emphasis should be placed on a careful analysis of what it is that different sides in the debate do and do not agree on and an increased willingness to engage in adversarial collaboration, in the spirit of a shared interest in furthering our understanding of animal behaviour.

Key-words: Animal cognition, Reasoning versus association, Morgan's Canon, Psychological continuity, Adversarial collaboration

Reasoning Versus Association in Animal Cognition: Current Controversies and Possible Ways Forward

The study of animal cognition is rife with controversy, and among the most long-standing and most intensely debated controversies in the field is the question to what extent the behaviour of non-human animals can be fully understood on the basis of purely associative principles, or whether some behaviours exhibited by animals necessitate the assumption of inferential capacities in animals that defy an associative explanation (Heyes, 2012; Penn & Povinelli, 2007). One of the most remarkable features of this controversy regarding the nature of animal cognition is its tenacity. Indeed, the origins of the debate can be traced back to Romanes (1882), Morgan (1894), and beyond (see Greenwood, this issue, for a historical overview). Yet, despite its long standing, it does not seem like we are making much progress towards consensus. Papers are continually being published suggesting cognitive capacities in non-human animals that defy an associative explanation, which keep being met with scepticism by people in the associative camp (indeed, the very fact that "camp" is a natural term to use in describing the situation is evocative of the degree to which this appears at times to be an ideological debate). It is true that some degree of controversy and disagreement is beneficial, as it provides fuel for scientific inquiry. However, for the present debate the continuing controversy seems to be spawning little genuine progress in terms of substantial accumulation of new, generally accepted, insights. As a result, the enduring controversy often seems more sterile than fruitful. Against this background, in late 2013 a special meeting was arranged by Tom Beckers and Jan De Houwer in Ghent to bring together people of different views in an effort to re-fertilise the debate and move beyond the current stalemate. The collection of papers assembled in the present special section represent (in part) the content and subsequent effects of that meeting. Here we outline a number of critical issues in furthering the debate, discuss how the papers in the special section contribute to that aim, and point to other developments that are likely to improve on the status quo.

Conceptual issues: History and the (in)adequacy of Morgan's Canon

The debate regarding the associative or more complex (that is, more cognitively rich; see Haselgrove, this issue) nature of animal cognition is strongly connected with the debate on the continuity between human and non-human animal cognition (see Shettleworth, 2012) and the status of Darwin's famous claim that "the difference in mind between man and the higher animals … is certainly one of degree and not of kind" (Darwin, 1871, p. 105). However, as Greenwood (this issue) points out, the connected debates over continuity/discontinuity of cognitive processes between humans and other animals and over the continuity/discontinuity of associative and inferential processes come with a great deal of (typically unacknowledged) historical baggage. In making this history more explicitly available Greenwood provides a timely caution against expecting entirely systematic and general answers in this field.

On one side of the debate, people have rightfully argued that while animals exhibit remarkably complex behaviour, such complexity does in itself not rule out that behaviour is determined by associative principles (Dickinson, 2012). Elementary associative principles, particularly when operating in combination, can yield surprisingly complex and rich behaviour (see Haselgrove, this issue). On the other side, people have rejected the idea that an associative explanation of behaviour should by default trump a more cognitive account for the same behaviour (see Hanus, this issue). While Morgan's Canon (Morgan, 1894; see Greenwood, this issue) indeed suggests that "lower" explanations for a given animal behaviour should be preferred over "higher" explanations, proper epistemological justification for Morgan's canon has been found wanting (Heyes, 2012). Moreover, the Canon is open to multiple interpretations and may support vastly different conclusions depending for instance on the scope of findings that is considered or the level of taxonomy at which it is applied – e.g., at the level of an entire class (e.g., the cognitive functioning of mammals), at the level of a specific order (e.g., the cognitive functioning of rodents), or at the level of a selected subset of a class (e.g., the cognitive functioning of all mammals other than humans). For example, it is now widely accepted that human cognition cannot fully be accounted for on the basis of purely associative principles. In the human causal learning field, discussion is now between those that support a single-process propositional view of human learning (e.g., Boddez, De Houwer, & Beckers, in press; De Houwer, 2009) and those that argue for a dual-process model (e.g., McLaren et al., 2014). Given that state of affairs, any comprehensive theory of learning that also aims to encompass human learning will need to include non-associative processes. But what is most parsimonious then: to assume that learning in non-human animals reflects simpler associative processes only, thus creating different theories for learning in different mammalian species, or assuming that supposedly more complex theories that explain human learning are also applicable to non-human mammals (thereby increasing the complexity of explanation for most mammalian species, but maintaining a single overall theoretical framework across mammalian species)? Greenwood (this issue) further argues that simplicity at one level of explanation (i.e., the biological level) need not be isomorphic with simplicity at another level of explanation (i.e., the psychological level). We would add that, considering arguments for a careful distinction between functional and cognitive levels of explanation within psychology (De Houwer, Hughes, & Barnes-Holmes, this issue; Hughes, De Houwer, & Perugini, in press), simplicity at the functional or behavioural level need not be isomorphic to simplicity at the cognitive or mechanistic level of explanation.

In summary, despite being so commonly invoked, Morgan's Canon does not appear to provide much help in choosing between different accounts of animal behaviour. Instead, it could be considered as an example of the sort of systematic general principle that Greenwood cautions us not to expect to work. But if we cannot rely on general principles to arbitrate between positions, then what should we do instead? The answer would appear to be deceptively simple: start without prior assumptions for or against particular classes of theories and look to developing decisive experiments (an idea echoed more or less explicitly in all the contributions to this special section). This is in no way a novel view (see Heyes, 2012), but its widespread acceptance should help breaking the stalemate created by incompatible default assumptions of principle. While it is important not to trivialise the difficulty in moving beyond established assumptions (see Hanus, this issue, for a discussion of some of these), the overview provided by Greenwood (this issue) may help in making it explicit that some entrenched positions owe more to history than to scientific principle.

Clarifying the terms of the debate.

To begin the process of developing decisive experiments without a-priori assumptions about general classes of explanation, it would be helpful to set out as clearly as possible what it is, and what it is not, that both sides of the debate disagree on. De Houwer and colleagues (this issue) begin their contribution by making a clear distinction between the behavioural or functional level of analysis (what it is that humans and non-human animals do and do not do) and the cognitive level of analysis (what difference in underlying cognitive capacities we infer from those behavioural differences). Clearly distinguishing between those levels should serve to illuminate the exact nature of the controversy. Sometimes, less than optimal research designs will cause disagreement about what it is that animals actually do, in functional terms, such as adapting their inspections of a food-containing hide in response to the potential presence or absence of a human observer (Taylor, Miller, & Gray, 2012, versus Dymond, Haselgrove, & McGregor, 2013; see Haselgrove, this issue) . More often, however, there is reasonable agreement about similarities and differences between human and non-human animals at the functional level, but strong disagreement about what those imply for the cognitive level – in particular when there are multiple candidate mechanisms in associative and inferential terms which are consistent with the same functional behaviours.

A second issue for clarification is what exactly is meant by "associative" or "rational" in the context of this debate. Implicit within the contributions by Hanus (this issue), as well as by Dwyer and Waldmann (this issue), is the idea that "associative" is often shorthand for a whole class of models, with various degrees of complexity. Even when each individual model is clearly specified on its own terms, considering the multiplicity of models and principles as a whole does not necessarily allow a simple unambiguous account for a given set of behaviours. Moreover, the diversification of contemporary associative theory can make it hard for those outside the field to know what principles are captured by this very general term (Hanus, this issue) and lead to the perception that associative theory can explain almost any possible pattern of behaviour (especially if applied in a post-hoc manner). Hanus rightfully argues that an associative explanation is to be preferred only if it is possible to come up with a clear and precise prediction of an experimental observation, not if it merely manages to posthoc explain anything. In such a case, a theory that provides precise prediction should be considered more parsimonious empirically. That principle ought to be applied with caution, however. While prediction is important in science (just like simplicity), the fact that one account predicted an effect and another only explained it post-hoc is not a logical proof that the first account is correct. Moreover, a similar argument about the excessive power of post-hoc explanation and opacity to outsiders can be made regarding cognitive or rational accounts of behaviour. Still, by placing the emphasis on the predictive value of different theoretical accounts, the challenge laid out in Hanus' contribution is for associative theory to be specified in a way that allows the expert and non-expert to know what it predicts.

In this light, the contribution by Haselgrove to this special section is particularly valuable in reviewing the current state of associative theory and providing an integrated description of a number of otherwise separate models and principles (as well as highlighting freely available resources which would facilitate simulating different models). This material is presented through a discussion of what needs to be done to overcome associative accounts, thereby setting a bar for anyone who wants to claim that a certain instance of non-human animal behaviour (or human behaviour, for that matter) defies an associative explanation.

One additional aspect of Haselgrove's contribution is to highlight something that associative accounts explicitly do not do – namely say anything about *how* two events are connected (see also Dwyer & Waldmann, this issue). The importance of the relationships between events is central to the contribution by De Houwer et al (this issue). De Houwer et al. discuss a functional-cognitive approach, and propose a new candidate distinction between human and non-human cognition that at the same time appears to allow for both some continuity and for some qualitative distinction between species, in terms of the arbitrariness of the types of relations that subjects can learn to respond to. To illustrate, humans can learn to relate almost any arbitrary stimuli in a multitude of ways, and derive novel relations as a result (e.g., to relate the word GLASS with the object glass and to relate the word VERRE to the object GLASS after it has been related to the word GLASS). We can learn and derive those types of relations like sameness, oppositeness, and others, for seemingly any arbitrary pair of stimuli. Non-humans, it appears, can also learn to respond to the relation between objects (e.g., learning to respond to the sameness of two stimuli presented in succession; e.g., Peña, Pitts, & Galizio, 2006) but only for stimuli that are non-arbitrarily related (e.g., physically identical).

Having considered (and rejected) the idea that broad general principles like Morgan's Canon could provide a means of arbitrating between competing accounts, and having looked at ways to clarify the exact issues in dispute, the next step is to consider the sort of empirical work which might help answer the questions that have been proposed.

Developing empirical tests

As we implied above, focusing on developing decisive experimental tests is simple only on the surface – as an answer to how to overcome the lack of progress in the field it begs the critical question by assuming we know what genuinely diagnostic tests would be. But do we? As we noted above, a multitude of research papers are being presented as demonstrating animal behaviour that cannot be explained associatively (and many arguing the opposite). We must assume that the authors of these papers believed that their work was diagnostic (at least at the time it was prepared for publication), and yet the field has remained in an effective stalemate. This brings us to the final question raised in the special section: what can be done to raise the probability that empirical tests will actually resolve critical issues? Or to put it another way, how do we ensure that one camp will take work performed by the other one seriously? We have already considered the ways in which clearly specifying the nature of different classes of theoretical accounts and resetting entrenched default positions could improve the quality of the debate. The same things should also enhance the impact of experimental work: For example, a better working knowledge of associative theory should assist in designing experiments aimed at disconfirming its predictions. However, there are many steps between theory and experimental design, and so improving the quality of theoretical knowledge will only go so far.

One possible way to make progress is to embrace the fact that researchers from different perspectives also bring different empirical and analytical expertise. Instead of performing research entirely "in house" and relegating the input from other perspectives to an after-the-fact analysis of the experimental work performed, the complementary expertise from different perspectives can be used in the development and conduct of the experiments themselves. That is, people from across the aisle can collaborate to develop a design that both parties agree could unambiguously decide between an associative and a rational/inferential account, and then perform that work. The contribution by Dwyer and Waldmann (this issue) represents an (incomplete) example of this process.

Given the divergence in perspectives, this is expecting a great deal from adversarial collaboration, as the exercise has proven to be far from trivial in execution, and it is unlikely to settle all debates (after all, no one researcher is a perfect representative of their "camp"). Notwithstanding these caveats, where such collaborations are feasible, they should prove to be a useful tool. This is partially because they instantiate the general ideas that we have already considered: The focus on empirical collaboration assumes that it is data rather than general principle that will decide the issue; the involvement of researchers from different perspectives mitigates against the unexamined influence of biased default assumptions; and joint experimental design requires explicit pre-experimental specification of the relevant predictions. Moreover, the very fact that the process as a whole is based on researchers from different perspectives working together means that they are taking each other seriously. We would also note that the times are clearly receptive to such adversarial collaborations, as they are on the rise in other fields of psychology (e.g., Matzke et al., 2015). Other emerging trends in experimental practice, in particular pre-registration and pre-experiment review (Chambers, Dienes, McIntosh, Rotshtein, & Willmes, 2014), also reflect key aspects of this approach through recognising the benefit of making the basis for theoretical claims explicit and open to external scrutiny prior to the conduct of experimental work[1].

While we have high hopes for progress based on collaboration and cooperation between different theoretical camps, it is also instructive to compare the broad general analyses of principle (from Greenwood, Hanus, Haselgrove, De Houwer et al.) against the more humble aims of the experimental work proposed by Dwyer and Waldmann. These authors are careful to point out that their proposed studies are not aimed at arbitrating between associative and inferential accounts in general, but instead are aimed at comparing one particular aspect of one inferential account (namely the influence of representing uncertainty within a causal model) against specified associative alternatives (and even this comparison is contingent on a number of simplifying assumptions). This is not a lack of ambition in Dwyer and Waldmann's proposal, but instead reflects the fact that empirical work is typically highly incremental, addressing focused comparisons between specified theoretical alternatives one at a time.

Cautious optimism about the ways forward

The six papers in this special section, and the meeting which inspired them, represent the efforts of people from a range of usually competing theoretical perspectives to explore together possible ways to reinvigorate the somewhat stalled progress in investigating the cognitive capacities of non-human animals. This is not an easy process – enculturation in a particular tradition imposes biases, some of which are explicit but many of which are implicit. Recognising these biases, and looking to move beyond the heuristics which reflect them to focusing on potentially discriminating empirical studies is a key step in this process. There needs to be a willingness to approach issues with an open mind and to try and find common ground, however limited it may be. Mutual recognition that many of our disagreements are situated at a cognitive level of analysis, which implies both sides of the debate rely on fallible inferences from data rather than on facts that are directly given, might be helpful in this regard. The rich set of principles represented by associative processes and the encompassing nature of inferential processes are both too broad to be amenable to direct falsification through a limited number of empirical studies. And nor should we expect them to be. Both a historical reflection, and the example of dual-process accounts in human cognitive psychology, caution against the expectation of simple and sweeping explanations applying generally across species and situations. We may have described this special section as "reasoning versus associations" but instead of seeking the (probably mythical) one true explanation of all animals behaviour, the route to genuine and lasting progress may well lie in looking incrementally for the best account of a multitude of specific behaviours. Overall, we feel there should be an emphasis on being collaborative rather than being adversarial – in the end, we all share the common goal of wanting to find out more about animal behaviour and its underlying mechanisms.

Footnote

[1] We would also acknowledge other recent developments in psychological research, including the emphasis on replication and meta analysis, detailed considerations of experimental power, and the limitations of classical statistics (along with the potential of Bayesian alternatives; see Lindsay, 2015). For new empirical work to have a lasting impact, it must also be reliable, and to the extent that these trends drive more reliable experimental work they will be as important for comparative psychology as they will be elsewhere.

Author Note

Tom Beckers, Department of Psychology, KU Leuven, Tiensestraat 102 box 3712, 3000 Leuven, Belgium; and Department of Psychology and Cognitive Science Center Amsterdam, University of Amsterdam, Postbus 15933, 1001 NK Amsterdam, the Netherlands; Jan De Houwer, Department of Psychology, Henri Dunantlaan 2, 9000 Gent, Belgium; Dominic M. Dwyer, School of Psychology, Cardiff University, Tower Building, 70 Park Place, Cardiff, CF10 3AT, UK.

The present paper and the special section of the *Journal of Comparative Psychology* that it is part of originated in a meeting on the nature of animal cognition organized by Tom Beckers and Jan De Houwer in Ghent in December 2013. Organization of that meeting and preparation of the special section were supported by Research Grant G.0766.11N and Scientific Research Network Grant 3WO00512 of the Research Foundation – Flanders (FWO), InterUniversity Attraction Pole Grant P7/33 of the Belgian Science Policy Office, KU Leuven Program Funding Grant PF/10/005 and Ghent University Methusalem Grant BOF09/01M00209.

Correspondence regarding the present paper can be addressed to any of its authors, at tom.beckers@kuleuven.be, jan.dehouwer@ugent.be, or dwyerdm@cardiff.ac.uk.

References

Boddez, Y., De Houwer, J., & Beckers, T. (in press). Inferential reasoning theory of causal learning: Towards a propositional account. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning*. Oxford, UK: Oxford University Press.

Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P. and Willmes, K. (2015). Registered reports: Realigning incentives in scientific publishing. *Cortex, 66,* A1-A2. doi: 10.1016/j.cortex.2015.03.022

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London, UK: John Murray.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37,* 1-20. doi:10.3758/LB.37.1.1

De Houwer, J., Hughes, S., & Barnes-Holmes, D. (in press). Associative learning as higher-order cognition: Learning in human and nonhuman animals from the perspective of propositional theories and Relational Frame Theory. *Journal of Comparative Psychology*.

Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society B, 367,* 2733-2742. doi: 10.1098/rstb.2012.0220

Dwyer, D. M., & Waldmann, M. R. (in press). Beyond the information (not) given: Representations of stimulus absence in rats (Rattus norvegicus). *Journal of Comparative Psychology*.

Dymond, S., Haselgrove, M., & McGregor, A. (2013). Clever crows or unbalanced birds? *Proceedings of the National Academy of Sciences, 110*, E336. doi: 10.1073/pnas.1218931110

Greenwood, J. D. (in press). All the way up or all the way down? Some historical reflections on theories of psychological continuity. *Journal of Comparative Psychology.* doi: 10.1037/a0039916

Hanus, D. (in press). Causal reasoning versus associative learning: A useful dichotomy or a strawman battle in comparative psychology? *Journal of Comparative Psychology*.

Haselgrove, M. (in press). Overcoming associative learning. *Journal of Comparative Psychology*.

Heyes, C. (2012). Simple minds: A qualified defence of associative learning. *Philosophical Transactions of the Royal Society B, 367,* 2695-2703. doi: 10.1098/rstb.2012.0217

Hughes, S., De Houwer, J., & Perugini, M. (in press). The functional-cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology*.

Lindsay, D. S. (2015). Replication in Psychological Science. *Psychological Science, 26,* 1827-1832. doi: 10.1177/0956797615616374

Matzke, D., Nieuwenhuis, S., van Rijn, H., Slagter, H. A., van der Molen, M. W., & Wagenmakers, E.-J. (2015). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. *Journal of Experimental Psychology: General, 144,* e1-e15. doi: 10.1037/xge0000038

McLaren, I. P., Forrest, C. L., McLaren, R. P., Jones, F. W., Aitken, M. R., & Mackintosh, N. J. (2014). Associations and propositions: The case for a dual-process account of learning in humans. *Neurobiology of Learning and Memory, 108,* 185-95. doi: 10.1016/j.nlm.2013.09.014

Morgan, C. L. (1894). *An introduction to comparative psychology*. London, UK: Walter Scott.

Peña, T., Pitts, R. C., & Galizio, M. (2006). Identity matching-to-sample with olfactory stimuli in rats. *Journal of the Experimental Analysis of Behavior, 85,* 203–221. doi: 10.1901/jeab.2006.111-04

Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology, 58,* 97-118. doi: 10.1146/annurev.psych.58.110405.085555

Romanes, G. J. (1882). *Animal intelligence*. London, UK: Kegan Paul, Trench, & Co.

Shettleworth, S. J. (2012). Modularity, comparative cognition and human uniqueness. *Philosophical Transactions of the Royal Society B, 367,* 2794–2802. doi: 10.1098/rstb.2012.0211

Taylor, A. H., Miller, R., & Gray, R. D. (2012). New Caledonian crows reason about hidden causal agents. *Proceedings of the National Academy of Sciences*, *109*, 16389-16391. doi: 10.1073/pnas.1208724109