Vol. 16 No. 2 (2025) 14–19
LETTERS ON EVOLUTIONARY BEHAVIORAL SCIENCE

OPINION

Beyond Anthropomorphism by Design: A Study-First Bridge for Evolutionary Psychology and Comparative Cognition

Kazuhiro Goto*

Sagami Women's University, 2-1-1 Bunkyo Minami-ku Sagamihara, Kanagawa, 252-0383, Japan

*Author for correspondence (kazuhiro.goto@gmail.com)

Evolutionary psychology and comparative cognition both pursue evolutionary accounts of cognition and behavior, yet they meet recurring interface problems that blunt strong inference: anthropomorphic projection, unvalidated human baselines, weak hypothesis construction, and conflation of proximate with ultimate explanations. I recast "comparative evolutionary psychology" as a study-first, field-scaling bridge rather than a disciplinary merger and propose a practical toolkit: design tasks calibrated to each species' sensory and motor capacities, establish constraint-matched human baselines, and present a priori predictions that pit specific adaptationist predictions against specified domain-general process models where they make divergent predictions. These predictions are adjudicated with diagnostic probes such as transfer to novel situations. Treating analysis as explicit model competition can raise the evidential bar, help reduce anthropomorphic bias, and better connect functional and mechanistic levels. These practices provide a principled path toward discovering genuine evolutionary continuities, convergences and divergences of cognition and behavior.

Keywords
evolutionary psychology, comparative cognition, anthropomorphism

---

Introduction

Evolutionary Psychology (EP) and comparative cognition (CC) are both fields that pursue evolutionary explanations of cognition and behavior. Although both emphasize the adaptive significance of behavior and the processes shaped by natural selection, they differ substantially in their primary targets and methodological approaches. EP typically infers the functions of present human mechanisms from hypothesized ancestral problems rather than reconstructing phylogenies directly. Instead, it regards present human psychological mechanisms as products of evolutionary adaptation and aims to reconstruct their functions and structures through a reverse-engineering approach (Oda, 2013). In contrast, researchers in CC—particularly those trained in the experimental analysis of behavior—seek to elucidate the diversity and commonalities of cognition and behavior across species by analyzing observable behavior in terms of stimulus control and reinforcement contingencies (Goto, 2019).

Recent efforts have sought to integrate these traditions into closer conversation to advance evolutionary accounts of cognition and behavior. Vonk and Shackelford (2013) proposed "comparative evolutionary psychology", combining cross-species comparisons with the tests of functional hypotheses. Despite a shared evolutionary grounding, the fields have followed distinct trajectories. CC has prioritized identifying both similarities and differences across species and often interprets them within ecological and phylogenetic contexts (Shettleworth, 2009). Although many CC researchers consider how particular traits may have emerged, the field has historically emphasized experimental demonstrations and mechanistic accounts, sometimes without explicit adaptationist modeling. EP, grounded in adaptationist reasoning, focuses primarily on the functional design of human psychological mechanisms (Tooby & Cosmides, 1992), yet often without systematic cross-species comparisons (Barrett et al., 2007), which can constrain broader evolutionary inference.

Vonk and Shackelford (2013) proposed a unified evolutionary program that treats proximate mechanisms and ultimate functions as complementary and bridges long-standing divides (field vs. lab; human- vs. animal-centered; behaviorist vs. cognitivist; nomothetic vs. idiographic) by using species-appropriate yet comparable tasks to analyze both similarities and differences (including convergence/divergence) across species. Progress has been partial: durable field-level integration remains limited, with most EP–CC bridging occurring at the article level rather than via shared training or standards. A small but clear set of study-level bridges has appeared—for example, unified cross-species self-control assays (e.g., MacLean et al., 2014), open-science consortia that standardize tasks across labs (e.g., ManyPrimates et al., 2019), and portable empathy/prosocial paradigms in rodents (Gachomba et al., 2024; with ongoing debate about social-contact alternatives in rat "helping"). Practical shifts include greater attention to species-calibrated design and uneven uptake of open-science practices, but two recurring gaps appear to impede synthesis—routine constraint-matched human baselines and genuinely commensurate cross-species analyses. As a study-first remedy, I specify standards that laboratories can, where feasible, implement now, allowing field-level integration to emerge cumulatively.

Finally, throughout this paper I treat proximate and ultimate explanations as complementary rather than competing. Although much work in EP centers on humans, comparative and developmental findings are increasingly

doi: 10.5178/lebs.2025.125
Received 03 September 2025.
Accepted 15 September 2025.

used as constraints; the present program formalizes how such constraints can be coupled with species-level experimentation. Following Tinbergen's framework, a mature account of behavior must specify both the mechanisms that generate it (e.g., stimulus control, reinforcement contingencies, sensory capacities) and the evolutionary pressures that favored those mechanisms. This stance helps avoid two reciprocal errors: reducing function to mechanism ("it's just reinforcement") and reducing mechanism to function ("it must be an adaptation"). The agenda developed below seeks to integrate the two levels by deriving functional predictions that constrain mechanistic models, while requiring mechanistic models to justify evolutionary interpretations by outperforming plausible non-evolutionary alternatives.

Commonalities between comparative cognition and evolutionary psychology

Although they differ in targets and methods, the fields share several important commonalities. First, both actively incorporate evolutionary theory when attempting to understand cognition and behavior; behaviors and mechanisms are positioned in terms of adaptive significance for survival and reproduction. Second, both address what Tinbergen (1963) called ultimate causes—the evolutionary functions and origins of behaviors. CC elucidates how cognitive traits emerged in evolutionary contexts through cross-species comparisons (Shettleworth, 2009), whereas EP derives functional hypotheses about psychological mechanisms from putative ancestral selection pressures (Tooby & Cosmides, 1992). Third, both deepen evolutionary understanding by comparing and analyzing extant species: CC maps commonalities and diversity across taxa, and EP sometimes leverages animal findings to illuminate human cognition.

Differences between comparative cognition and evolutionary psychology

Despite shared foundations, the two traditions diverge in how they generate and evaluate evidence (Table 1). EP tends to reason top-down: it treats present-day human psychological systems as solutions to ancestral adaptive problems and evaluates reverse-engineered functional models by how well they predict current patterns (Confer et al., 2010). CC works bottom-up: it establishes, via tight experimental control, which variables actually govern behavior—including stimulus control and reinforcement contingencies—and then asks how those controls generalize across species, even for phenomena sometimes described as self-reflective (Goto, 2012).

A second difference concerns scope. EP concentrates on human cognition, often probing putatively human-specific designs, whereas CC samples a wider range of taxa to map both continuity and divergence in cognition. In doing so, comparative studies routinely consider species' sensory capacities, body plans, and motor repertoires when interpreting how tasks are solved.

Finally, the two fields emphasize different kinds of explanatory payoff. EP prioritizes functional fit—does a hypothesized design solve a plausible ancestral problem?—while CC prioritizes experimentally demonstrated control—what proximal variables produce the behavior, and do those variables align across species? Taken together, these emphases are complementary: functional hypotheses can guide what to test comparatively, and comparative analyses can constrain which functional stories remain viable. This complementarity motivates the integrative agenda developed in the next section.

Comparative evolutionary psychology revisited: A bridge, not a merger

Building on this common ground and divergence, Vonk and Shackelford (2013) proposed *comparative evolutionary psychology*, an approach that couples the formulation of evolutionary hypotheses about psychological mechanisms with empirical tests grounded in cross-species comparisons. The aim is to deepen understanding of cognitive and behavioral design by comparing multiple species, including humans, and by moving beyond simple presence–absence demonstrations of particular abilities. The emphasis instead is on the convergence and diversity of solutions to adaptive problems across both closely and distantly related species: asking not only whether a trait exists, but why it evolved and under which selection pressures.

Crucially, the framework aspires to integrate functional (ultimate) and mechanistic (proximate) explanations. In practice, however, broad disciplinary integration has been limited; progress has tended to occur at the article level. Accordingly, in this paper I treat comparative evolutionary psychology as a bridge at the EP–CC interface rather than a call for merger. I operationalize that bridge via study-level practices that make results portable between fields. In principle, this approach can help overcome human-centered biases, anthropomorphic interpretations, and

Table 1. Overview of evolutionary psychology and comparative cognition.

	Evolutionary Psychology (EP)	Comparative Cognition (CC)
Species	Primarily human	Primarily nonhuman animals
Mechanistic emphasis	Domain-specific functional systems; domain-general processes as constraints/rivals	Domain-general learning/mechanisms; species/ecology-specific functions
Typical evidence	Human experiments/surveys; cross-cultural tests; translational nonhuman evidence when homologous (e.g., hormones, neural signals); computational/functional models	Cross-species comparative experiments; diagnostic probes; computational models

Note. Both fields study mixes of domain-specific and domain-general mechanisms; the table reflects predominant emphasis, not exclusivity.

oversimplified narratives by systematically probing relationships among species, behaviors, and environments, while preserving each field's distinct aims.

The proposal takes a study-first approach that scales to the field level. At the study level, researchers can, where feasible, (i) calibrate tasks to each species' biology, (ii) collect human baselines under matched constraints, and (iii) preregister rival functional and mechanistic predictions plus diagnostic probes. At the field level, many such studies cumulate into comparative meta-analyses that weigh model families across taxa and contexts. Integration is thus not an all-or-nothing merger but an adoptable design/analysis standard.

Methodological challenges at the comparative cognition-evolutionary psychology interface

This article adopts a CC standpoint. The four challenges below are not equally prevalent across fields: anthropomorphism, unvalidated human baselines, and biology-calibrated design are more salient in CC, whereas weak hypothesis construction/post hoc explanation is field-general (Table 2). I treat them as interface challenges because they recur where evolutionary claims draw on comparative evidence and where proximate–ultimate slippage can arise on either side. The remedies proposed are study-level practices intended to make results portable across the EP–CC interface.

1. Anthropomorphism

At the EP–CC interface, anthropomorphism manifests as over-attributing human-like states to nonhumans in CC and as human-centric assumptions in EP task design/interpretation; an a priori model rivalry can help keep both in check when rival accounts are operationalized and penalized comparably. Both fields are prone to anthropomorphic projection—the risk of attributing human-like emotions, intentions, or cognition to nonhuman animals without sufficient empirical justification. For example, human visual assessments have long misclassified many bird species as sexually monochromatic, whereas modeling avian visual systems (including ultraviolet sensitivity) reveals widespread dichromatism (Eaton, 2005). This illustrates how human sensory assumptions can systematically distort biological interpretation. Likewise, when animals display helping, it is tempting to infer empathy or prosociality (Ben-Ami Bartal et al., 2011; Sato et al., 2015), yet such behavior may be explained by simpler contingency structures; careful analysis is required to avoid over-attributing internal states (Schwartz et al., 2017).

A practical antidote is to formalize rival process accounts (often dubbed *killjoy* hypotheses) a priori and adjudicate among them with quantitative fits to behavior. In rodent helping, for example, one can pit a reinforcement-learning model (treating social contact as an appetitive outcome with state- and history-dependent values) against an empathy-based account that predicts state-dependent generalization across contexts. The empathy account should survive only if it explains transfer, extinction, outcome devaluation (e.g., reduce the value or availability of social contact, deliver it non-contingently, or mask distress cues), and contingency degradation better than the learning model at comparable complexity.

In putative metacognition tasks, before inferring access to confidence states, first probe—and where appropriate, retain—parsimonious accounts by testing whether "information-seeking" is controlled by identifiable external stimuli and/or shaped by reinforcement history for responding to those stimuli; if these factors account for the pattern, a stimulus-control explanation suffices (Hataji & Goto, 2024).

When EP uses nonhuman results to support human functional claims, require a priori *killjoy* predictions and

Table 2. Evolutionary psychology–comparative cognition interface problems and study-level remedies.

Interface problem	Salience	Why it matters at the interface	What to do
Anthropomorphism	CC > EP	Nonhuman→human inferences can over-attribute states/functions; EP designs can import human centric defaults.	Specify a priori *"killjoy"* alternatives; add diagnostic probes (transfer, devaluation, contingency degradation); retain mentalistic accounts only if they beat penalized simpler models.
Unvalidated human baselines	CC > EP	Cross-species anchors and EP-only claims both mislead when the "human" reference is assumed, WEIRD, or instruction-dependent.	Collect constraint-matched human baselines; estimate strategy mixtures.
Biology-calibrated design	CC > EP	Differences in what can be sensed or done can look like differences in evolved function.	Calibrate to species' sensory/motor capacities; translate apparatus by function not form; include within-species decoys/controls and between-species functional replications.
Weak hypotheses	discipline-wide	Adaptationist or mechanistic narratives can be tailored ex post; failures become uninformative.	Preregister hypothesis, use diagnostic manipulations; fit formal process models; use equivalence tests/Bayes factors for "no-difference" claims; favor parsimony via penalized selection.

Note. Salience codes (e.g., comparative cognition (CC) > evolutionary psychology (EP)) are heuristic field impressions, not prevalence estimates. *Discipline-wide* = commonly observed across both EP and CC.

diagnostic probes; retain anthropomorphic interpretations only if they outperform penalized alternatives.

2. Unvalidated human baseline

In CC this appears when animal data are judged against an assumed human "default"; in EP it appears when convenient tasks stand in for evolved function—both require constraint-matched human baselines before cross-species or functional claims. A recurrent, under-acknowledged problem is the unvalidated human baseline: researchers assume "typical" human behavior within a paradigm and project that assumption onto nonhumans without first testing humans under identical perceptual, instructional, and motor constraints. The Guesser–Knower paradigm illustrates this point. Although it is often presumed that humans reliably prefer the knowledgeable agent, Gagliardi et al. (1995) showed that adults frequently followed the Guesser when doing so yielded greater reward than following the Knower, implying that performance tracked salient cues and payoff structure—i.e., stimulus control—rather than knowledge attribution. This undermines the very baseline used to evaluate nonhuman performance (cf. Povinelli et al., 1990). A second illustration comes from transitive inference. Classic work in pigeons showed above-chance choices on novel non-adjacent test pairs after adjacent-pair training (e.g., BD after training on A+B−, B+C−, C+D−, D+E−), establishing that nonhuman animals can succeed on transitive inference under standard conditions (von Fersen et al., 1991). Yet adult humans can perform at chance on the BD test when explicit awareness of the ordered list is curtailed, succeeding only on easier contrasts like BE (Frank et al., 2005)—i.e., transfer can fail without explicit structure, cautioning against treating human performance as the default benchmark.

A parallel caution comes from chimpanzee working memory. Inoue and Matsuzawa (2007) reported that young chimpanzees outperformed human adults on a masked-numeral memory task. However, when human participants received adequate practice on the same task, their performance met or exceeded that of the chimpanzee, indicating that the apparent species gap reflected training history rather than a qualitative cognitive advantage (Cook & Wilson, 2010).

Tasks intended to probe physical causal understanding often build in a single "correct" response, yet both chimpanzees and adult humans show systematic biases that do not track causal necessity. For example, in trap-tube variants adults unnecessarily insert the stick from the end farthest from the reward, and in trap-table variants they avoid the side with an (ineffective) hole; changing distances can even reverse preferences (Silva et al., 2005). These patterns indicate that performance can be governed by spatial heuristics or motor habits rather than causal insight. Because humans succumb to the same biases, chimpanzee "failures" cannot be taken as evidence against causal reasoning per se; instead, the paradigms are not diagnostic. Before drawing species differences, establish human baselines under matched instructions, viewing, and motor demands, and include controls that pit proximity-based strategies against causal predictions (Povinelli, 2000; Povinelli & Dunphy-Lelii, 2001).

In many comparative contexts, a validated human baseline (strategy-mix estimation plus stress tests such as payoff flips, time pressure, or sensory masking) is important both for cross-species anchors and for EP-only studies to keep functional claims from collapsing into untested heuristics.

3. Biology-calibrated paradigm design and validation

Although especially salient in CC, biology-calibration also governs EP claims: without species-appropriate instructions, perceptual channels, motor demands, and payoffs, task success or failure can masquerade as functional design. Design tasks with careful attention to each species' sensory systems, body plan, and motor repertoire—the constructive counterpart to avoiding anthropomorphism. First, sensory calibration: experiments are often built so that experimenters can readily perceive the manipulated variables, yet animals may exploit cues outside human perception (e.g., ultraviolet patterning in birds; ultrasonic signals in rodents). Accounting for such nonhuman-accessible channels is essential to ensure that the task manipulates the intended variable rather than an unnoticed one. Second, functional equivalence: complete sameness of task across species is impossible; the realistic goal is for nominally "the same task" to manipulate comparable functional variables, not merely to reuse a look-alike apparatus. Third, cross-validation: include within-species controls that dissociate perceptual salience from functional relevance (e.g., decoy cues that attract attention but are causally inert), and conduct between-species replications that preserve function while altering form through apparatus translation.

Piloting with humans under matched constraints (limited instructions; identical viewing and motor demands) is indispensable: if humans fail or switch strategies once verbal scaffolds or demonstrations are removed, the paradigm is unlikely to be diagnostic for nonhumans. As the chimpanzee memory case illustrates, practice and instructional scaffolds can shift human performance from chance to above-chimp levels on identical procedures; without matched exposure, cross-species contrasts risk misattributing experience effects to evolved differences (Cook & Wilson, 2010).

Finally, frame analyses to compare competing mechanisms—stimulus control, reinforcement history, heuristic rules, and model-based control—rather than to confirm a favored construct. Biology-calibrated validation raises the evidential bar and increases power to detect genuine continuities and meaningful divergences in evolutionary solutions. Taken together, adopting these study-level standards enables cumulative cross-species meta-analyses that compare model families across taxa.

Constraint-matched tasks (instructions, perceptual channel, response modality, payoff structure) mitigate task-format confounds distinct from WEIRD (Western, Educated, Industrialized, Rich, Democratic; Henrich et al., 2010) sampling and prevent task constraints from masquerading as evidence of evolved function, and make nonhuman-to-human inferences defensible.

4. Weak hypothesis construction and post hoc explanations

Both traditions risk narrative fit without a priori stakes; a priori rival predictions, diagnostic manipulations, and penalized model comparison make successes informative and failures equally model-diagnostic. A discipline-wide vulnerability of psychology is ease with which adaptationist stories can be tailored to fit observed behavior without yielding a priori, falsifiable predictions (cf. Gould & Lewontin, 1979). In CC, striking performances are sometimes reported without functional predictions that would separate convergent evolution from plasticity or artifact; in EP, broad claims about putative adaptations can outpace evidence when rival domain-general accounts are not entertained seriously. This pattern echoes Farrar and Ostojic's diagnosis of confirmation-seeking designs, ambiguous constructs, and post hoc narratives that blunt strong inference (Farrar & Ostojic, 2019).

A practical remedy in both fields is to enforce model rivalry and tighten inference. Specify a priori predictions, and adjudicate them with diagnostic probes (e.g., transfer tests to novel situations). Fit formal process models (e.g., reinforcement learning) and use penalized model selection so that simpler, domain-general accounts prevail whenever they fit as well. When the theoretical claim concerns an absence or practically small effect, use equivalence tests (Lakens, 2017) or Bayes factors rather than defaulting to non-significance (Farrar et al., 2023; Mulder & Wagenmakers, 2016). A constraint-matched human baseline remains essential: test humans under the same perceptual, instructional, and motor constraints before exporting a paradigm, so the benchmark is evidence, not assumption (Cook & Wilson, 2010; Frank et al., 2005; Gagliardi et al., 1995; cf. Povinelli et al., 1990). Finally, ensure construct clarity and task validity so that paradigms manipulate the theorized variable rather than an incidental cue (Farrar & Ostojic, 2019).

Conclusion

EP and CC share a commitment to explaining cognition and behavior in evolutionary terms—and they face overlapping vulnerabilities: anthropomorphic projection, unvalidated human baselines, weak hypothesis construction, and slippage between proximate and ultimate levels. The framework advanced here addresses these issues by (i) designing species-calibrated tasks that guard against human-centric blind spots, (ii) establishing human baselines under matched constraints, and (iii) specifying a priori rival predictions that link functional claims to mechanistic models, with adjudication via diagnostic probes and parsimonious, learning-based *killjoy* alternatives.

Importantly, *killjoy* is not joyless: such explanations are a central and productive approach in biology, demonstrating how complex patterns can emerge from simple processes (Shettleworth, 2010). Treating them as serious competitors strengthens comparative inference rather than diminishing it.

Building a shared methodological framework is not a call to add "more species" to human-centered hypotheses or "more function" to mechanistic demonstrations. It is a commitment to diagnostic design plus explicit model rivalry. Where preregistered (or explicitly stated) rival predictions succeed across species, we can provide principled signals of continuity; conversely, informative failures can indicate divergence. The field advances when findings arbitrate among narratives, not when narratives are retrofitted to findings.

Author contribution

KG conceptualized and wrote the manuscript.

References

Barrett, L., Henzi, P., & Rendall, D. (2007). Social brains, simple minds: Does social complexity really require cognitive complexity? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1480), 561–575. https://doi.org/10.1098/rstb.2006.1995

Ben-Ami Bartal, I., Decety, J., & Mason, P. (2011). Empathy and pro-social behavior in rats. *Science*, *334*(6061), 1427–1430. https://doi.org/10.1126/science.1210789

Confer, J. C., Easton, J. A., Fleischman, D. S., Goetz, C. D., Lewis, D. M. G., Perilloux, C., & Buss, D. M. (2010). Evolutionary psychology: Controversies, questions, prospects, and limitations. *American Psychologist*, *65*(2), 110–126. https://doi.org/10.1037/a0018413

Cook, P., & Wilson, M. (2010). Do young chimpanzees have extraordinary working memory? *Psychonomic Bulletin & Review*, *17*(4), 599–600. https://doi.org/10.3758/PBR.17.4.599

Eaton, M. D. (2005). Human vision fails to distinguish widespread sexual dichromatism among sexually "monochromatic" birds. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 10942–10946. https://doi.org/10.1073/pnas.0501891102

Farrar, B. G., & Ostojic, L. (2019). The illusion of science in comparative cognition. PsyArXiv. https://doi.org/10.31234/osf.io/hduyx

Farrar, B. G., Vernouillet, A., Garcia-Pelegrin, E., Legg, E. W., Brecht, K. F., Lambert, P. J., Elsherif, M., Francis, S., O'Neill, L., Clayton, N. S., & Ostojić, L. (2023). Reporting and interpreting non-significant results in animal cognition research. *PeerJ*, *11*, Article e14963. https://doi.org/10.7717/peerj.14963

Frank, M. J., Rudy, J. W., Levy, W. B., & O'Reilly, R. C. (2005). When logic fails: Implicit transitive inference in humans. *Memory & Cognition*, *33*(4), 742–750. https://doi.org/10.3758/BF03195340

Gachomba, M. J. M., Esteve-Agraz, J., & Márquez, C. (2024). Prosocial behaviors in rodents. *Neuroscience and Biobehavioral Reviews*, *163*, Article 105776. https://doi.org/10.1016/j.neubiorev.2024.105776

Gagliardi, J. L., Kirkpatrick-Steger, K. K., Thomas, J., Allen, G. J., & Blumberg, M. S. (1995). Seeing and knowing: Knowledge attribution versus stimulus control in adult humans (*Homo sapiens*). *Journal of Comparative Psychology*, *109*(2), 107–114. https://doi.org/10.1037/0735-7036.109.2.107

Goto, K. (2012). How can anthropomorphism be incorporated into the comparative cognition studies? *Japanese Journal of Animal Psychology*, *62*(1), 49–57. https://doi.org/10.2502/janip.62.1.8

Goto, K. (2019). Hikaku ninchi [Comparative cognition].

In The Japanese Association for Behavior Analysis (Ed.), *Kōdō bunsekigaku jiten* (pp. 392–395). Maruzen Publishing.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Hataji, Y., & Goto, K. (2024). Information-seeking in mice (*Mus musculus*) during visual discrimination: study using a distractor elimination paradigm. *Animal Cognition*, *27*(1), Article 81. https://doi.org/10.1007/s10071-024-01920-3

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, *17*(23), R1004–R1005. https://doi.org/10.1016/j.cub.2007.10.027

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J., Burkart, J. M., Call, J., Cantlon, J. F., Cheke, L. G., … Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), E2140–E2148. https://doi.org/10.1073/pnas.1323533111

Many Primates, Altschul, D. M., Beran, M. J., Bohn, M., Call, J., DeTroy, S., Duguid, S. J., Egelkamp, C. L., Fichtel, C., Fischer, J., Flessert, M., Hanus, D., Haun, D. B. M., Haux, L. M., Hernandez-Aguilar, R. A., Herrmann, E., Hopper, L. M., Joly, M., Kano, F., … Watzek, J. (2019). Establishing an infrastructure for collaboration in primate cognition research. *PLOS ONE*, *14*(10), Article e0223675. https://doi.org/10.1371/journal.pone.0223675

Mulder, J., & Wagenmakers, E.-J. (2016). Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments." *Journal of Mathematical Psychology*, *72*, 1–5. https://doi.org/10.1016/j.jmp.2016.01.002

Oda, R. (2013). Shinka to ningen kōdō [Evolution and human behavior]. In H. Ihobe & R. Oda (Eds.), *Kokoro to kōdō no shinka o saguru: Ningen kōdō shinkagaku nyūmon* (pp. 1–35). Asakura Publishing.

Povinelli, D. J. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works*. Oxford University Press.

Povinelli, D. J., & Dunphy-Lelii, S. (2001). Do chimpanzees seek explanations? Preliminary comparative investigations. *Canadian Journal of Experimental Psychology*, *55*(2), 185–193. https://doi.org/10.1037/h0087365

Povinelli, D. J., Nelson, K. E., & Boysen, S. T. (1990). Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *104*(3), 203–210. https://doi.org/10.1037/0735-7036.104.3.203

Sato, N., Tan, L., Tate, K., & Okada, M. (2015). Rats demonstrate helping behavior toward a soaked conspecific. *Animal Cognition*, *18*(5), 1039–1047. https://doi.org/10.1007/s10071-015-0872-2

Schwartz, L. P., Silberberg, A., Casey, A. H., Kearns, D. N., & Slotnick, B. (2017). Does a rat release a soaked conspecific due to empathy? *Animal Cognition*, *20*(2), 299–308. https://doi.org/10.1007/s10071-016-1052-8

Shettleworth, S. J. (2009). The evolution of comparative cognition: Is the snark still a boojum? *Behavioural Processes*, *80*(3), 210–217. https://doi.org/10.1016/j.beproc.2008.09.001

Shettleworth, S. J. (2010). Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences*, *14*(11), 477–481. https://doi.org/10.1016/j.tics.2010.07.002

Silva, F. J., Page, D. M., & Silva, K. M. (2005). Methodological-conceptual problems in the study of chimpanzees' folk physics: How studies with adult humans can help. *Learning & Behavior*, *33*(1), 47–58. https://doi.org/10.3758/BF03196049

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift für Tierpsychologie*, *20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). Oxford University Press.

von Fersen, L., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*(3), 334–341. https://doi.org/10.1037/0097-7403.17.3.334

Vonk, J., & Shackelford, T. K. (2013). An introduction to comparative evolutionary psychology. *Evolutionary Psychology*, *11*(3), 459–469. https://doi.org/10.1177/147470491301100301