**Animal morality and epistemic risks[[1]](#footnote-1)**

Nicolas Delon ([ndelon@ncf.edu](mailto:ndelon@ncf.edu))

New College of Florida

**Abstract.** There is growing evidence that many animals engage in what appears to be morally charged behavior. I draw on epistemological considerations to assess the risks of making claims about animal morality under uncertainty. Being wrong about animals, when our beliefs can affect how we treat them, is morally risky. An important thesis about the relation between values and the epistemic domain is *Inductive Risk*. Nonepistemic considerations, including moral values, bear on the epistemic status of our beliefs, hypotheses, and interpretations of the evidence. I argue for stricter evidentiary standards when choosing null hypotheses and accepting hypotheses relating to animal morality. Should we always default to overattribution (err on the side of epistemic caution)? I argue that we should not assume that the moral consequences of error regarding animal morality follow an epistemic precautionary principle (i.e., when in doubt, overattribute). Instead, a comprehensive assessment of moral costs is required. However, psychological evidence suggests that the claims about animal morality pass the test.

1. **Introduction**

A growing literature documents animal “proto-morality”, spearheaded early on by primatologist Frans de Waal and colleagues (de Waal 2006; 2009). Many primates exhibit “building blocks of morality”: empathy, consolation, conflict resolution, cooperation, fairness (or inequity aversion) (Flack and de Waal 2000). While animals lack full-blown morality, they manifest behavior that is to some degree moral—genuinely prosocial and other-regarding. Ethologist Marc Bekoff and philosopher Jessica Pierce have argued that many species can follow moral norms (of empathy, fairness, cooperation, and mutual help), but that such norms are species-specific: there is human morality, wolf morality, rat morality, and so on. Besides a blossoming empirical literature, philosophical work on animal morality has also exploded (e.g., Andrews and Gruen 2014; Behdadi 2020; Monsó and Andrews, forthcoming; Ferrin 2019; Fitzpatrick 2017; Monsó 2015; 2017; Monsó et al. 2018; Monsó and Wrage 2021; Rowlands 2012; Shupe 2021; Vincent et al. 2019).

I draw on epistemological considerations to assess the risks of making claims about animal morality under uncertainty. I appeal specifically to *arguments from inductive risk* (Douglas 2000; 2009; Rudner 1953) to argue that moral considerations bear on the level of evidence required for our claims to be epistemically justified—whether we are in the business of accepting or rejecting hypotheses or that of forming credences (degrees of belief) about hypotheses. Being wrong about animals, in contexts in which our beliefs can affect how we treat them, is morally risky. While the usual focus of such arguments is on *failures* to attribute certain mental properties to other animals, I argue that the reverse, attributing properties that they lack, also deserves scrutiny. The question is whether we should accept stricter evidentiary standards regarding animal morality.

In a context of widespread speciesism, exploitation, and ‘dementalization’ of other animals, adopting a liberal attitude toward what hypotheses to accept regarding animal minds seems warranted. Then, the thought goes, we should

1. Either apply stricter standards to hypotheses to the effect that humans possess a feature that animals of certain lack (or possess it to a much greater degree), where the null hypothesis is that there is no significant difference between humans and such animals,
2. Or relax the standards required to reject the null hypothesis, where the hypothesis tested is that such animals possess the feature.

Intuitively, the potential harms associated with not recognizing the moral capacities of animals if they have them exceed whatever harms might be associated with recognizing them if they lack them. And given the low probability that learning more about animals’ mental and moral lives would make us more susceptible to mistreating them the costs of being wrong do not seem egregious, even granting inductive risk.

How we see animals matters, but we should not assume that anything we may learn about them will necessarily lead us to treat them better. If animals have moral status and it depends at least partly on their mental capacities, then inductive risk tells us to avoid under-attribution rather than over-attribution. (We also have purely epistemic reasons to avoid over-attribution, but it’s a separate question.) However, while denying animals some traits may reinforce speciesist attitudes and exploitation, we overlook the possibility that some cases of over-attribution could be harmful too. If directly or indirectly contributing to the perception of animals as moral poses moral risks, our claims about animal morality require a higher evidentiary threshold. If, for example, some animals meet the requirements for responsibility (Ferrin 2019) and deeming and holding responsible imposes potential costs on them, then we have moral reason to ensure that our claims meet that threshold.

Intuitively, however, claims about animal morality do not seemof that sort. Contrast two kinds of claims:

*Negative characteristics*

* Chickens are stupid
* Pigs live fully in the present
* Other animals are gross
* Pit-bulls are killers
* Foxes are cunning
* Animals act purely out of instinct

*Positive characteristics*

* Elephants mourn their dead
* Rats experience empathy
* Chimpanzees use tools
* Octopuses recognize individual faces
* Orcas cooperate in hunts
* Crows engage in social learning

Attributions of moral characteristics, it seems, fall under the ‘positive’ umbrella. They are not degrading or disrespectful, they do not objectify animals or reinforce prejudices about their inferiority. If anything, the opposite. Moreover, moral subjects are not just moral *patients*, and it’s plausible they deserve distinctive protections and respect (Rowlands, 2012: 248-254; Monsó et al., 2018). So whence the worry?

I argue that the inductive risk associated with mistakenly attributing morality to other animals is not negligible, by which I mean that it should be assessed instead of assuming that attributing morality necessarily benefits them. As we’ll see, it is unlikely to be bad in most contexts, but the point is this is not something we know *ex ante*. Recent work on methodology in animal cognition focuses on the risks associated with failing to ascribe certain cognitive capacities to other animals (Andrews and Huss 2014; Birch 2017; Mikhalevich 2015). Traditionally, the risk of over-attribution has been seen as worse than that of under-attribution. We should reject the presumption of priority of false negatives over false positives, for reasons both scientific and ethical. Both are errors, and the former can have high ethical costs. But there are also risks to attributing capacities that animals lack. Because our evaluations of behavior tend to be associated with presumptions of accountability, evaluating animal morality risks percolating into our assumptions about what we should expect of animals. And as the stakes go up, so should out epistemic standards.

The paper proceeds as follows. §2 sets up the argument from inductive risk. §3 applies it to the animal morality debate. §4 discusses whether denying that animals are moral is harmful and proposes a guiding principle of Inductive Risk Calibration. §5 reviews psychological literature suggesting that a lack of clear evidence that attributions of moral abilities to other animals adversely affects our treatment of them.

1. **Inductive risk**

What is inductive risk and how does it help characterize the epistemological challenges we face in animal cognition?

Jonathan Birch writes,

Animal welfare scientists face an acute version of the problem of “inductive risk” … since they must choose whether to affirm or reject uncertain hypotheses about the mental capacities of animals, knowing that their decisions may hold significant consequences for animal welfare. (2018: 1027)

Inductive risk occurs in contexts of uncertainty, when affirming or rejecting a hypothesis bears a risk of error. It takes various forms depending on the types of errors one is considering. A type I error (or false positive) involves incorrectly rejecting a true null hypothesis; a type II error (or false negative) involves a failure to reject a false null hypothesis.

In animal cognition research, scientists have traditionally used some interpretation of Lloyd Morgan’s controversial Canon (Morgan 1894; Sober 2000, 2005; Allen-Hermanson 2005; Fitzpatrick 2008; Andrews and Huss 2014). The 19th century comparative psychologist was a critic of the method of ‘anecdotal anthropomorphism’ of Darwin and Romanes. According to Morgan’s Canon, we should, absent sufficient evidence to the contrary, prefer mental explanations of behavior in the simplest terms available. In statistical methods, this translates into a preference for skeptical null hypothesis combined with a preference for type II errors (Mikhalevich [2015] characterizes this as “underattribution bias”, a combination of an error-rate asymmetry, i.e., a preference for type II errors, witha bias toward parsimonious null hypotheses, i.e., that posit fewer or simpler mechanisms). Thus, the null hypothesis is usually the absence of the mental phenomenon of interest in the target species. False positives, or type I errors, are errors of overattribution, whereas false negatives, or type II errors, are errors of underattribution.

Epistemic caution is a scientific virtue. However, inductive risk means that not all errors are equal. Errors of underattribution, in high stakes contexts, may be worse than errors of overattribution. Thus, regarding sentience, Birch argues that we should *reverse* the traditional direction of epistemic caution and default to taking false negatives more seriously than false positives (2017; 2018). Underattribution is worse than overattribution because of the costs of falsely denying that, say, pigs or chickens can feel pain or pleasure. One might then be tempted to employ a generalized principle of the following form (Birch 2018, 1030):

For any mental state *M*, underattributing *M* creates far more serious risks of negative animal welfare outcomes than overattributing *M*. So, in the context of advising on animal welfare policy, an animal welfare scientist should affirm the hypothesis that organisms of species *S* have *M* whenever there is credible scientific evidence that organisms of *S* have *M*, even if that evidence is inconclusive and subject to continuing debate.

The principle is roughly “the inverse of Morgan’s canon”: *when in doubt, err on the side of overattributing mental states*.

Are there contexts where we should default to the standard priority of epistemic caution, not just contexts where the stakes are low or epistemically pure, but when they flip—where it would be worse for animal welfare to overattribute than to underattribute. As Birch notes,

… when there are clear policy applications in view … when animal cognition research directly informs animal welfare regulations, a special context is created in which erroneously affirming a false mental state attribution may be a less serious error, all things considered, than failing to affirm a true attribution. (1028).

It is not obvious that the moral consequences of error, in the case of animal morality, recommend prioritizing the avoidance of errors of overattribution. The context is that of scientific communication affecting our perception of animals but also concrete regulations likely to affect the way we treat animals who possess moral abilities, from farms to zoos, from sanctuaries to wildlife reserves.

Let’s take the cue from Birch himself, who notes that, although the anti-canon “may seem attractive at first sight, it oversimplifies a complex issue … the relative seriousness of over- and underattribution depends on the species, mental state, and animal welfare intervention in question.” (1030). Birch’s example is that of cognitive enrichment and the impact on welfare of its implementation based on insufficient evidence. Cognitive enrichment interventions are a subset of environmental enrichment designed to enhance the welfare of captive or domesticated animals by creating opportunities to express certain behaviors, to exercise certain capacities, and to gain control over aspects of their environment. But crucially,

What counts as a cognitive enrichment for a particular animal depends on the cognitive capacities of that animal, and there must be a close match between cognitive ability and environmental design. A challenge that is so difficult as to induce stress or anxiety is not an enrichment; a challenge so easy as to induce boredom or apathy is not enrichment either (1031)

Animal cognition research is directly relevant to designing effective interventions. This is where inductive risk enters: risks of overattribution must be considered too, lest we design interventions that either fall short of or exceed animals’ abilities in ways that might compromise their welfare. The main risk consists in “jumping too quickly” to a conclusion that such and such enhancement is conducive to positive welfare. We could cause unnecessary stress, there could be opportunity costs, or we might prevent stress when it could facilitate learning (1035).

Because inductive risk cuts both ways, Birch recommends a principle centered on animal welfare to guide hypothesis testing. Inductive risk tells us that how much scientific evidence we need to accept, reject, or fail to reject hypotheses depends on the level of non-epistemic risks associated with the risk of error. The impact of research on animal welfare policy is a source of non-epistemic risks, which must be accounted for.

Thus, as Birch shows, a precautionary principle does not always yield the correct verdict regarding welfare: “it seems appropriate to require more than the mere existence of credible scientific evidence that an organism possesses a cognitive ability before recommending an enrichment premised on its possession of that ability.” (1032) The reason why the enrichment context requires a “different burden of proof” than that of minimizing pain is because the stakes are different: underattribution is higher stake in the case of pain than in the case of enrichment, whereas overattribution raises few if any risk in the case of pain but does raise some risks in the case of enrichment.

1. **Inductive risk and animal morality**

The animal morality debate illustrates what Birch calls the danger of “evaluative anthropomorphism”, where we project our own values on to animals. It is tempting to think that, if seeing human beings as incapable of moral emotions leads us to treat them poorly, the same is true of animals. But we just noted evidence to the contrary. We cannot simply assume that attributing more complex mental capacities necessarily has the same welfare implications across species. What we need, first, is a “richer theoretical understanding of the nature and causes of good psychological welfare in animals.” (Birch, 2018: 1036) Absent an account of the welfare implications of (mis)attributing moral abilities to animals, we may not assume that the precautionary principle applies (for a start, see Monsó et al., 2018). I return to this question in the next section. For now, let’s return to hypothesis testing.

If we opt for a liberal null hypothesis concerning, say, capacities for caring, epistemic caution can lead us to prefer false negatives over false positives and therefore to fail to reject the null hypothesis. If our evidence base is simply credible but not very robust, we may be tempted to assume, in doubt, that all Fs are Gs—that pigs are generally capable of caring for conspecifics. Now combine this assumption with an assumption about the positive contribution to welfare of the freedom to exercise such a capacity, and you get the conclusion that inductive risk warrants *treating* pigs *as* possessing a moral ability. Again: when in doubt, err on the side of overattribution.

The problem is that our null hypotheses are themselves informed by the available evidence. If such evidence is already biased toward overattribution, we risk building a weak body of merely credible evidence and recommending on its basis interventions that may not match the animals’ capacities or, regardless, might misrepresent their contribution to welfare. Inductive risk does not so much call for precaution as for a proper assessment of the impact of research on welfare.

Birch suggests an alternative, “expected welfare maximization framework” based on the following principle:

the burden of proof for affirming a mental state attribution, in the context of advising on animal welfare policy, should be set so as to maximize the expected total welfare of the nonhuman animals affected by the policy. (2018: 1032)[[2]](#footnote-2)

If the expected disvalue of being wrong exceeds the expected disvalue of not attributing moral abilities, we should act *as if* animals lack them or wait until implementing interventions that presuppose such capacities. In contrast, the precautionary principle would have told us that, because of uncertainty and given some evidence, we should act *as if* the target species possess such capacities (Birch, 2017; Sebo, 2018). In contrast, the application of the expected value framework is sensitive to the “specific advisory context” in which researchers inform policy decisions. The respective risks of over- or underattribution are relative to that context (Birch, 2018: 1034).

Likewise, in the context of animal morality the “welfare consequences of error” deserve careful empirical scrutiny. My view is not that we should not investigate animal morality because of inductive risk. Rather, because the “causal path” between attributions of capacities and welfare is empirical, more research is needed, as the saying goes. The practical question is not whether we should revert the error-rate asymmetry and abide by Morgan’s Canon, but how we should construct our hypotheses and research programs.

Many argue that exclusively epistemic factors (evidence and background knowledge) should feed into our decisions (Mikhalevich, 2015; Andrews and Huss, 2014). For instance, Andrews and Huss argue (*pace* Morgan’s Canon):

There are often cases where pragmatic or moral concerns might justify counting a particular hypothesis as the null hypothesis. [For example, in drug trials ‘‘drug x is ineffective’’, in criminal jurisprudence ‘‘innocent until proven guilty’’] *But when our concerns are purely epistemic, as they presumably are in the case of animal cognition*, it’s less clear why either the skeptical or optimistic hypothesis should get preferential treatment from the outset. (Andrews and Huss 2014, 721; emphasis added)

But because of inductive risk, some pragmatic (including moral) factors should also influence our hypotheses at the margin. Between two equally conservative/liberal hypotheses, pick the one that minimizes expected welfare costs according to Birch’s principle. The empirical question (see §5) is whether there is evidence one way or another concerning animal morality.

Kristin Andrews (2011) is right that, in some contexts, type II errors can constitute a fundamental error (falsely denying a property) worse than type I errors (falsely attributing a property). Why? Because in such contexts the preference for type II errors affects future research by shaping research programs and restricting the scope of legitimate questions. Combined with skeptical (as opposed to liberal) null hypotheses, the preference for type II errors can indeed be epistemically problematic. Research is path-dependent, and the error-rate asymmetry leads to the construction of a body of evidence biased against animals, which in turns informs hypotheses that we think we can reasonably entertain and frameworks we use to investigate animal minds and behavior.

However, by the same token, type I errors can inform research programs in ways that have been overlooked, of which cognitive enrichment provided an illustration. Admittedly, such errors are less likely to restrict the scope of legitimate questions but are likely nonetheless to shape research programs and inform the construction of hypotheses. So, yes, there may also be risks to failing to attribute moral capacities to a target species, but there should be *some* credible evidence that failing to do so risks compromising welfare. We would have to think that a sizeable marginal improvement in how we treat the target species turns on that question.

By way of illustration, suppose we designed policies for designing sanctuary infrastructure and protocols for socialization in the target species based on our current understanding of their capacities (or lack thereof). Depending on whether we believe they are moral creatures, we might make different choices, and such choices likely affect the welfare of the animals. But how our choices affect welfare depends on how well they match the capacities that the animals do possess. And so again, we must consider inductive risk. As argued, we cannot simply assume that an abundance of charity (a bias towards overattribution) involves what’s best for the species. What if, for instance, engaging in moral behaviors causes stimulates animals in ways that cause them anxiety or assuages stress that would benefit for them? What if they involve opportunity costs preventing animals from engaging in more rewarding behaviors? What if environments that would be good for moral creatures are bad for nonmoral creatures and we incorrectly believe that the target species is moral? These are costs of error at various stages of inquiry. Some have to do with scientific inquiry itself, others with science-based policy or the implementation of policy. They all suggest that the link between animal morality and welfare is more complex than it seems. The evidence is overwhelming that we treat many animals poorly because we deny them mental characteristics they do possess. The question is whether this is always true. The answer has two parts: does morality contribute do welfare, and does seeing animals as moral benefit or harm them?

1. **Morality and welfare**

Applying Birch’s expected welfare maximization framework, we can ask what specific risk of error is worse: incorrectly failing to attribute or misattributing moral abilities. The respective magnitude of these costs is then what informs our evidentiary standards.

Even if we correctly believed that some animals are moral creatures, we should be wary of the inductive risks posed by the new hypotheses our correct beliefs generate—people are bad at internalizing the nuance of scientific information; popular science reporting tends to overhype and exaggerate the actual results of scientific studies; correct statements facilitate hypotheses that are on much shakier ground; etc. We might be correct that some animals are moral but incorrect about the contribution of their abilities to their well-being, or we might be correct about both but tend to overgeneralize or extend the import of such beliefs beyond their appropriate scope. We should only give animals the benefit of the doubt that is required for underattribution not to have expected negative welfare outcomes. A plausible principle would be:

*Inductive risk calibration:* Attribute as much as but no more than an animal’s well-being requires, given the causal path between the distribution of epistemic risks and the distribution of welfare outcomes*.*

Let’s consider the first part of the answer. Susana Monsó et al. (2018) focus on the hypothetical example of a sow, whom they call Sustitia2 (to distinguish her from Sustitia1, a sow comparable in all respects except that she is not a moral subject).

What makes her different from Sustitia1 is that Sustitia2 does not just suffer due to *her own life conditions*, she is also concerned with the well-being of the sows and piglets in her environment. … Sustitia2 is characterised by the possession of a mechanism in her brain that ensures that whenever she witnesses a conspecific in distress, she too undergoes a form of distress that (1) is *intentionally directed* at the distress of the conspecific, and (2) has an *urge to engage in affiliative behaviour* built into it. (291)

Sustitia2 is a ‘moral subject’, who can act for moral reasons (has moral motivations), but she is not a moral agent (Rowlands, 2012).[[3]](#footnote-3) They then argue that, if we deny that she is capable of empathy and care, we are more likely to deny her the conditions to exercise her capability or to deprive her of the capability itself. If moral capabilities are constituents of flourishing, our denial could harm her.

The argument turns on a crucial assumption—that the possession or the exercise of moral capabilities is a *necessary* component of flourishing for certain species. Monsó et al. appeal to Nussbaum’s capabilities (2006) approach to contend that moral abilities are basic capabilities that necessarily constitute flourishing for creatures who can possess them (Monsó et al., 2018: 296). Also implied is the precautionary idea that we should treat Sustitia2 *as if* she had moral capabilities even if she might not. She could be harmed if we believe she is not capable of moral behavior, either because we will deprive her of opportunities to engage in such behavior, or because depriving her of such opportunities is necessarily associated with other harms such as bodily mutilations, isolation, or confinement.

What do we gain by calling such behavior *moral*? What matters is for the sow to be able to engage in affiliative activities with her piglets and other pigs, not whether such behavior counts as moral. We need not call it moral to recognize that it is essential to flourishing. If, moreover, believing that the sow has moral abilities leads us to falsely believe that she has further needs or preferences, our treatment may not align with her best interests, as we saw with cognitive enrichment. If, finally, calling the sow moral facilitates, albeit illicitly, ascriptions of responsibility, we might be inclined to expect too much of her, to blame her for not properly taking care of her piglets or to be more favorably disposed towards ‘better’ mothers. These are speculative worries; the point is we should weigh them instead of presuming that calling animals moral can only benefit them.

Calibrating the balance of inductive risk in setting epistemic standards for hypothesis testing is key, as per the above principle. This involves, among other things, resisting exaggerations that might appear charitable or good for the cause. Exaggerations run afoul of certain epistemic standards but those standards themselves are sensitive to non-epistemic factors (Douglas 2009). Writing about hype in scientific communication, Kristen Intemann writes,

whether an instance of communication constitutes hype depends on two sorts of value judgments. First, … a value judgment about what are the most important goals for communicating with the audience in a particular context. Exaggeration is inappropriate when it hinders the goals most important to that audience, such as predictive relevance or accuracy, for the sake of excitement, reassurance, or interest. Second, exaggeration is inappropriate to the extent to which the implicit and explicit inferences made are insufficiently supported by the existing evidence, given the risks of being wrong. (Intemann 2020, 9)

Does the literature on animal morality impose problematic risks on its audience by hindering “important goals of communication” or inviting “risky inferences that do not meet appropriate evidentiary standards”? The answer depends on our goals and what risks we are willing to accept. This brings into relief the context-sensitivity of Birch’s framework: different standards are appropriate depending on whether research seeks to inform policy to issue specific guidelines, ‘raise awareness,’ generate testable hypotheses, question assumptions about human superiority, and so forth. The balance of inductive risk depends on these different aspects of the epistemic and practical context.

Suppose the audience encompasses scholars and laypersons, as such relatively unlikely to exert influence on policy. What matters is what risks are involved in having them act upon the findings of inquiry if such findings are wrong, how we can expect their behavior to be changed by new studies.[[4]](#footnote-4) One place to look at is the literature on the psychology of human-animal relationships.

1. **Psychology**

My argument from inductive risk rests on the possibility that attributions of morality could have negative welfare consequences for animals in certain contexts. The specific argument follows from the general position that animal cognition research is not value-free. However, its upshot turns on empirical matters. In this section, I review some evidence that defuses the initial worry. However, the empirical resolution is precisely the kind of work that an application of Inductive Risk Calibration requires.

Our moral judgments are sensitive to a range of new information about animals’ capacity for experience (e.g., capacity for pain and emotions), agency (e.g., problem solving, tool use, coordination), and benevolence (e.g., they are gentle, caring, or peaceful). Could seeing animals as moral inadvertently reinforce stereotypes of animals as threats or ‘pest’? Some studies suggest that attributing morally laden capacities to animals can affect our attitudes. Jared Piazza and colleagues (2014) have found that perception of harmfulness (having a harmful as opposed to benevolent disposition, relative to human welfare) has a negative effect on attributions of moral standing, independently of their sentience or intelligence. Because “perceiving an animal as having a benevolent disposition enhances people’s moral consideration for that animal, which is likely to promote better treatment of it” (121), seeing animals are moral or immoral could have unforeseen consequences.

Not all moral animals play nice—predation, aggression, and callousness are pervasive. Our perception of predators could change if we saw them as moral subjects. We might see chimpanzees, dolphins, and orcas as sometimes immoral. Our attitudes to coyotes and wolves, already considered a nuisance by farmers and the US Fish and Wildlife Service, could further deteriorate. Such moral costs must be part of inductive risk assessments. Many animals are already stigmatized, shunned or despised, if not blamed, for behavior for which they are not morally responsible ([Kasperbauer](https://link.springer.com/article/10.1007/s10539-015-9478-y), 2017; Piazza et al. 2014). There is no evidence that adding reactive attitudes to the package would benefit animals. Directly, seeing animals as moral could increase our tendency to hold negative reactive attitudes toward them. Indirectly, it could lead to punishment (Shupe, 2021) or simply inadequate handling.

Across a series of four studies, Stefan Leach and colleagues (2020) found that the characteristics that most strongly influenced judgements about the moral treatment of other animals, specifically eating meat, were

the capacity to feel secondary emotions (e.g., love), understand morality (e.g., sharing food with others), empathize with others (e.g., feeling others pain), form social bonds (e.g., looking for deceased family members), and be harmed e.g., feel pain). (8)

Animals who were perceived to have such characteristics were “consistently perceived to be more wrong to eat than animals that had other capacities.” (ibid.) The studies suggest that ‘experience’ (capacity for feeling) appears to be driving attributions much more significantly, whereas the literature on ‘mind perception’ has often taken ‘agency’ (e.g., capacity for thinking, planning, and making decisions) as an independent source of moral status (Sytsma and Machery 2012). This suggests that people’s judgments about moral status are more likely to be sensitive to new information that pertains to experience. Crucially, animal’s morality “cues greater perceptions of experience” (5), because it is typically operationalized in terms of emotions and caring or empathetic behavior.

These findings inform our expected value assessment and defuse our initial worry. The studies corroborate the *positive* influence of perceptions of secondary emotions (Demoulin et al. 2004; Leyens et al., 2001) and morality and social connections (Haslam and Loughnan 2014; Piazza et al. 2014). Thus, it seems we may err on the side of overattribution when it comes to morality, at least insofar as we construe it in experiential terms (e.g., empathy or care vs. planning or knowledge). Yet the facts about animal behavior that inform moral judgement tend to be fine-grained. (Leach et al., 2020: 8) Not just anything we can learn will shift moral judgments in the same direction. Perceptions of agency *without experience* do not account for much of the variance, and the *kind* of agency matters a great deal.

The studies align with previous work suggesting that attributing characteristics associated with benevolence or harmfulness induce differential concern people for animals (Goodwin and Benforado 2015; Piazza et al. 2014; Weisman et al. 2017). For instance, sharing food with others had one of the largest effects on judgments about the wrongness of eating an animal. On the other hand, perceptions of *harmful agency* tend to have the opposite effect. Piazza et al. (2014) have argued that having a “cruel nature”, a dispositional tendency to inflict harm, is a third factor along experience and agency and drives lower attributions of moral status than to animals who are perceived to have a more benign disposition. The mere disposition, or “character trait,” is what drives the judgment independently of the actual ability of the animal to exercise it (“cruel” animals who pose no actual threat still have decreased moral status).

Albeit not moral character proper, harmfulness pertains to the animal morality debate (Goodwin 2015a, 42). Indeed, the opposite dispositions (benevolence and such) are readily associated with morality in the literature! Meanwhile, harmfulness is less justifiable as a normative source of moral status than either agency or experience; it is extrinsic and dependent on human interests and biases. Goodwin notes, “The fact that participants did find harmfulness relevant seems to represent an encroachment of ordinary principles of social cognition into a domain in which they may not justifiably apply” (Goodwin 2015b, 919).

**Conclusion**

In sum, the evidence suggests that perceptions of morality are not negatively associated with judgments of moral standing. Even if the animal morality literature was misguided, it doesn’t seem like it would lead us to treat animals more poorly. But another upshot is that this is not something we knew *ex ante*. The argument from inductive risk forced us to assess the whole context, including the audience’s psychology (Inductive Risk Calibration). The same principle could yield different conclusions based on the policies that animal morality leads to in zoos, biomedical research, and conservation. But what will count as acceptable evidence depends partly on nonepistemic facts and sometimes attributing more complex abilities to animals does not benefit them.

**References**

(To be completed)

1. A longer version of this paper also discusses Moral Encroachment, a related but distinct thesis on the relation between values and epistemic standards according to which whether a particular epistemic state has a particular epistemic status (e.g., justification, knowledge, plausibility, evidence) depends on some of its moral features. [↑](#footnote-ref-1)
2. One might reply, we don’t know how to evaluate expected welfare without an account of the contribution of moral abilities to welfare. But maybe we don’t need a precise account. Standard methods (whether subjective, objective, or hybrid) for measuring animal welfare will apply to moral abilities just like they apply to other characteristics. My claim is simply that, when conducting the assessment, we should not neglect the possibility that attributions of moral abilities might make a negative contribution in certain contexts. [↑](#footnote-ref-2)
3. “While we cannot praise Sustitia2 for her behaviour, given that she lacks moral responsibility, we should take into consideration that, whenever she comforts others in distress, she is doing so on the basis of a motivation that implies experiencing as bad something that is bad (namely, the conspecific’s distress), and so she is feeling how she *should* feel, given the circumstances.” (Monsó et al., 2018: 297) [↑](#footnote-ref-3)
4. Two caveats. (i) The risks involved in having the audience act on true findings are also worth considering but let’s bracket them here. (ii) I’m assuming an endogenous link between attitudes and behavior, but this is only a toy model of how to approach inductive risk in this context. Note, though, that denying this link doesn’t help the argument that seeing animals as moral will benefit them. [↑](#footnote-ref-4)