

Word Counts: Abstract – 92; Main Text – 7390 ; References – 407; Entire Text – 8281

Author's Response

Bifocal Stance Theory: An effort to broaden, extend, and clarify

Robert Jagiello^a robert.jagiello@anthro.ox.ac.uk

Cecilia Heyes^b cecilia.heyes@all-souls.ox.ac.uk

Harvey Whitehouse^a harvey.whitehouse@anthro.ox.ac.uk

^aCentre for the Study of Social Cohesion, School of Anthropology and Museum Ethnography, University of Oxford, Oxford, OX2 6PE, UK

^bAll Souls College and Department of Experimental Psychology, University of Oxford, Oxford OX1 4AL, UK

Abstract

The Bifocal Stance Theory (BST) of cultural evolution has prompted a wide ranging discussion with broadly three aims: to apply the theory to novel contexts; to extend the conceptual framework; to offer critical feedback on various aspects of the theory. We first discuss BST's relevance to the diverse range of topics which emerged from the commentaries, followed by a consideration of how our framework can be supplemented by and compared to other theories. Lastly, the criticisms that were raised by a subset of commentaries allow us to clarify parts of our theory.

1. Introduction

The Bifocal Stance Theory (BST) has stimulated a rich set of commentaries, bearing testimony to the productivity of our framework in advancing our understanding of human learning behaviour and cultural evolution. Many commentaries argued that BST has a broader range of application than even

we envisaged, for example to topics as diverse as language, music, education, pedagogy, psychopathology, and non-human animals. Many also suggested additional factors to be considered in the theory, or compared it to a competing framework thus generating testable predictions for future research. Our account of BST has also led to some criticisms that we seek to address. Our response is thus divided into three main parts dealing respectively with broadening the application of our theory, extending it, and responding to critique (Table 1).

	Contribution	Commentator
Broadening the applications of BST	BST in nonhuman animals	Whiten
	BST in music research	Loui & Margulis
	BST in science and education	Watve & Watve
	Evolutionary origins of BST	Veit & Browning
	BST and psychopathology	Campbell & Fonagy
	BST and Confucian philosophy	Lai & Stapleton
	BST and child-led teaching	Bazhydai & Karadag
	Agent-based models of BST	Leibo, Köster, Vezhnevets, Guzmán, Agapiou & Sunehag
	BST in language evolution	Scharinger & Erfurth
Extending BST	BST and threat sensitivity	Samore & Fessler
	The role of culture in shaping BST	Fong, Nielsen & Legare
	BST, within-culture variability and developmental milestones	Clegg, Wen & Rawlings
		Puttre & Corriveau

BST and Construal Level Theory (CLT)	Kalkstein & Trope
BST and social relationships	Thomas, Radkani & Hung
BST and the specificity, riskiness and complexity of goal-directed action	Brown & Pain
Cost-benefit calculations in BST	Vélez, Wu & Cushman
Repetition, goals and habits in BST	Perez
<hr/>	
Addressing the critics	
Social learning is more complex than BST	Packer & Cole
All rewards are instrumental	Zentall
Extrapolation from single learning instances	Buskell & Charbonneau
Environmental scaffolds drive social learning	Nonaka
Most of culture is learned via the instrumental stance	Dubourg, Fitouchi & Baumard
Most of magic is instrumental	Ze Hong
Relevance explains social learning	Altıñok, Tatone, Király, Heintz & Gergely
Arbitrary action is instrumental	Fong, Nielsen & Legare

Table 1. Overview of commentaries based on their general aim to provide applications, extensions or critiques of BST.

2. Broadening the Applications of BST

First and foremost, BST is a theory of social learning and hence it concerns every aspect of human life in which individuals acquire information through interaction with others. As the commentaries make clear, the explanatory value of our framework spans a vast range of different fields of inquiry, giving insight into the psychological mechanisms, evolutionary origins and meta-contributions of BST.

2.1. Underlying cognition

Campbell & Fonagy, argue that some symptoms in psychopathology could result from failure to read social cues correctly, producing too much or too little trust in social interactions, impairing the ability to switch flexibly between stances during learning interactions. This is a potentially fruitful idea to explore although it is important to remember that our framework proposes that the stances “lie in the eye of the beholder” and so, at least in principle, there is no “right” or “wrong” interpretation of any given action sequence. Nevertheless, we agree that deviation from normal patterns of stance adoption in a group could impair communication and learning observed across many different conditions. Most importantly, Campbell and Fonagy propose that non-standard adoption of ritual or instrumental stances may be due to deficits in mentalizing abilities. Our article has explored different types of cognitive structures underlying BST and remains open to this possibility. Nonetheless, in its simplest form, we propose that instrumental and ritual stances operate on domain-general processes of attention, motivation and learning. Rather than appealing to domain-specific mechanism, abnormal stance switching could be due to inadequate input in early development. This would be consistent with the looking glass paradigm we propose (see section 5), in which instrumental learning may reap social rewards while cues that are indicative of conventional learning are paired with inanimate rewards. A reversal of reward structure or complete lack of consistency in the way rewards are obtained may cause individuals to routinely adopt “the wrong” stance during ontogeny. Future research will be able to disentangle the contributions of rich, specialised processes, such as mentalising, from those of leaner, domain-general processes.

An important step in that direction has been taken by **Leibo, Köster, Vezhnevets, Guzmán, Agapiou & Sunehag** who consider how an agent-based modelling approach to BST could help clarify the various ways in which patterns of stability and innovation in social learning may be generated based on domain-general and non-deliberative cognitive processes. This approach marks a particularly exciting new direction for our framework, which to date has mainly focused on empirical research while leaving untapped the range of possibilities that agent-based models are capable of generating. They propose a domain-general learning model in which agents develop the tendency to punish norm deviations. Explaining the emergence of such a norm-punishing property is straightforward once the relevant scaffolding of promoting punishing behaviour is in place, but a pressing question would be to determine the origins of such a scaffold: how did the first generation of norm-punishers come into

being? Applying BST to agent-based modelling addresses not only how the stances can emerge independently, but also potentially pinpoints the factors that promote generic learners to switch between them in dynamic fashion. The successful implementation of such a model would thus weaken the case for a purely deliberative account of BST, shedding light on plausible candidate mechanisms mediating differential learning modes, as well as discussing the range of factors that may have given rise to ritual and instrumental stances in the first place.

2.2. Evolutionary origins and human uniqueness

Veit & Browning propose a feedback loop including cooperative foraging, and reliance on others, as origin of the stances. This however raises the question of why other species that also rely on group membership for improved survival have not developed a similar propensity to engage in ritual copying. Additionally, other factors could have played a pivotal role in the inception of this co-evolutionary interplay between the stances. It could be that the advent of hunting tools (such as spears and projectiles) increased costs of inter-group conflict which necessitated stronger social cohesion and improved group delineation via ritual learning. The link between war and ritual has been established in previous research (Sosis, Kress, & Boster, 2007; Whitehouse, 2021) and human warfare likely scaled up with instrumental learning, refining the weaponry involved in conflict, which in turn increases the importance of cementing group membership via accurate ritual transmission. Such a feedback loop is analogous to what Veit and Browning have suggested, but perhaps points to a different driver.

Indeed, a similar explanation is explored by **Samore & Fessler**, who discuss the link between adherence to tradition (via the ritual stance) to improved outcomes in threat situations (caused by environmental hazards, like pathogens or outgroup violence) through increased group cohesion. We agree with the authors that our framework is informative and possesses explanatory potential in elucidating the role of tradition in coping strategies in response to threat. To further investigate this link, future cross-cultural research might employ a combination of threat priming scenarios (e.g. risk of contamination, outgroup hostility, natural disaster, etc.) together with a subsequent learning phase in which subjects are asked to imitate actions from members of their own group. Would elevated threat salience promote the adoption of a ritual stance during subsequent learning episodes?

Whiten highlights another way of addressing the question of origins by considering an alternative possibility—that the ritual stance is not uniquely human but could be phylogenetically more ancient

than previously assumed. We agree with his suggestion that our initial formulation of BST—proposing that it mainly explains cultural patterns seen in humans—is not quite ambitious enough, as our framework might come with implications for explaining learning more widely within the tree of life. As Whiten astutely observes, this raises the question as to whether the observed conformism is produced by the motivation to fit into the new social environment or whether it hinges on the implicit assumption that other group members possess an informational advantage—in which case the copying behaviour is designed to achieve the best instrumental outcome. One way to probe this question might be to observe how ostracism affects these learning strategies. A particularly efficient proxy for social concern might be the individual’s rank within the group. If non-human primates make use of bifocal stances in the way they copy actions from others, then a low ranking or peripheral individual should act to secure group membership via conformism.

2.3. BST beyond the action domain

We equally agree with **Whiten** that BST should be applied to a wider range of human and non-human cultural phenomena, such as language and song. Comparative research that is mindful of the different types of perceived action (Figure 3 in the target article) will be needed to provide clarity as to how these phenomena fit into our framework. Even though songs may be attributed to an attainable goal, the means by which this end-state is achieved is causally opaque in irremediable ways. From the perspective of natural causation, there is no physical reason as to why specific sounds should be arranged in any particular way to create music. We therefore expect musical practices to persist with high accuracy and we find it particularly intriguing that bird and whale songs have not only been found to serve as coordination devices within specific populations but also to exhibit remarkable stability over time, both characteristics which are in line with the predictions of BST.

Indeed, music, song and language are topics that lend themselves to further investigation in light of BST, as demonstrated by **Loui & Margulis**, who apply BST to the domain of music by proposing an experimental design which aims to disambiguate the different pathways proposed by our cultural action framework, while also discussing how our framework can account for the evolution of music more generally. We find their argument that the innovative spirit of music composers comes from an instrumental stance, while other more socially focused uses of music activate the ritual stance, intriguing. Nonetheless, as we note within the context of quasi-instrumental practices as well as in our

response below (section 4.1.), the presence of an end-goal does not automatically result in the adoption of an instrumental stance.

Like Loui and Margulis, **Scharinger & Erfurth** apply BST's core arguments about the fidelity of imitated action to a non-action domain, proposing that the instrumental stance could drive language innovation while a speaker's motivation to affiliate with their group could promote linguistic stability via the ritual stance. BST is primarily a theory which attempts to explain imitated action, therefore we generally approach its applications outside of the action domain with some caution. For instance, the assumption that regular past tense forms accomplish their communicative goals via knowable causal pathways (hence prompting an instrumental stance) while irregular forms are causally opaque in irresolvable ways (because they do not undergo the rule application of the *-ed* suffix, hence prompting a ritual stance), is currently beyond the scope of our theoretical framework. Neither regular nor irregular verb forms can be potentially explained via natural causation in the way they reach their communicative goals. Nonetheless we agree that this is an interesting new angle which invites novel ways of theorizing.

An analogous approach would be to consider BST's role in the wider context of religion (Whitehouse, 2011; Whitehouse, 2012). For instance, do minimally counterintuitive concepts (MCI; Boyer, 2002; Nyhof & Barrett, 2001) serve a similar affiliative function as action sequences that are perceived as irretrievably opaque? Future research might establish whether socially and goal-driven motivations modulate the fidelity with which certain story elements and narratives are retold, thus further exploring the reach of BST's explanatory potential.

Lai & Stapleton's literary and philosophical approach makes use of the *Analects* (a historical text describing the lives of early Confucians) as a case study for BST, finding many parallels between our framework and the Confucian approach to ethico-social learning. This also prompts novel questions about the role of tailoring – the ability to enact a tradition accurately but with slight deviations as to accommodate the context in which it takes place – and about the relationship between the stances. We do sympathize with the authors' aim of drawing attention to the importance of tailoring, which places goal-focus in the context of ritual practice. Tailoring is a concept which aligns well with the quasi-instrumental practices we discuss in our article. We propose that learners switch flexibly between ritual and instrumental stances based on the weighing of cues, so while the presence of a goal might render an instrumental stance more likely (resulting in relatively more deviations during

ritual enactment) we maintain that given the importance of other cues in stance adoption, the presence of a goal does not deterministically send a learner into an instrumental learning mode (as discussed below in section 4.1.). Moreover, we applaud the author's application of BST to this case study as it further underlines our framework's versatility and reach in accounting for a wide variety of cultural phenomena.

Although we anticipate considerable flexibility in the interplay of the stances, we are inclined to resist Lai and Stapleton's suggestion that a varifocal lens would be a better metaphor than a bifocal lens. It may well be that flexible, culturally-shaped deployment of the ritual and instrumental stances yields a continuum of copying fidelity; the area between highly innovative and highly conventional behaviour may be thickly populated with intermediate cases. However, the hypothesis that this output continuum is generated by two distinct clusters of psychological properties – two stances in a bifocal relationship – has the virtue of coherence with existing evidence and of testability in future work. By contrast, it is hard to see how a varifocal account of BST could be adequately tested since any result that fails to conform to the predicted patterns of social learning associated with ritual and instrumental stances could simply be 'explained away' by being assigned to an intermediate middle ground.

2.4. BST in education

Watve & Watve's discussion of how our framework could lead to a better understanding, and ultimately optimization, of the practices and norms in academia and education harmonizes well with Whitehouse's (2021) argument that BST can help us better understand and address the general divide between the sciences and humanities (the "two cultures problem"). Here, the instrumental stance appears to motivate fields dominated by scientific thinking, focusing on causal transparency and attainable end-goals, while the ritual stance is more pervasive in the arts and humanities, which place greater emphasis on irremediably opaque discourse, inviting exegetical interpretation in much the same ways as rituals and artworks. We are equally intrigued by the prospect of investigating how the two stances may be adopted *within* a field, where for instance some aspects of science might place emphasis on faithful replication to preserve the current state of knowledge, which in turn could impede rates of innovation. Fostering an instrumental stance in the classroom via focus on end-states might be an exciting new objective to explore within the context of academia and education.

3. Extending BST

3.1. Competing hypotheses

Thomas, Radkani & Hung argue that non-instrumental actions communicate social relationships rather than delineating group membership. For example, they argue that bowing or kissing are poor markers of group identity because they are so ubiquitous. We are unconvinced, however, by the examples chosen. Cross-cultural research indicates great variability in how different groups kiss, with some cultures not engaging in the practice at all (Jankowiak, Volsche, & Garcia, 2015). Variations in kissing behaviour appear to be among the stereotypes used to delineate group identities, for example differentiating British, French, and Italian styles of greeting or taking leave.

Further, as the authors note, their examples describe behaviour that is asymmetric. This marks an important departure from BST as our framework aims to explain differences in copying fidelity in intergenerational and historical transmission of cultural practices. In other words, when a social learner observes a model and is prompted to replicate a cultural practice, why do they imitate some aspects more accurately than others? It appears that behaviours which do not inspire imitation as a direct consequence of observation, such as the mentioned example of bowing, are less relevant to BST. For instance, there is evidence that kissing serves important signalling purposes in sexual selection (Hughes, Harrison, & Gallup Jr, 2007) and we agree that it is unlikely to be a behaviour that requires a ritual stance in order to persist as it is anchored in biological proximate mechanisms—in much the same way as other forms of intimacy would not require ritual learning even though their immediate purpose may not be accessible to the actor or recipient of the action.

Accordingly, BST does not claim to account for all observed human behaviours but rather attempts to account for the psychological motivations that mediate instances of social learning within the cultural domain. The question then as to how social learners decide whether a non-instrumental action is about either rituals or relationships may have a rather simple answer: if the social learner is in a social situation that encourages or requires the replication of non-instrumental action, they may readily adopt the ritual stance based on detecting conventional cues. If the situation is not about learning but simply observing or online coordination of action, they may draw a plethora of different inferences based on the nature of the interaction, including assumptions about relationships such as described in this commentary. Moreover, this basic distinction in how a situation is framed can be investigated empirically by manipulating the contexts in which actions are observed as well as the implicit

expectations that are directed at the observing party in the social interaction. Do social learners copy rituals more accurately even if they did not expect to be asked to do so?

A theory which more closely competes with BST within the domain of social learning is Construal Level Theory (CLT) as discussed by **Kalkstein & Trope**. They propose that the higher the psychological closeness to a model the greater the motivation to copy observed actions with higher fidelity. Conversely, greater psychological distance promotes focus on a goal and thus higher abstraction of the steps that lead up to it. This model is rooted in the assumption that higher distance increases processing load, thus prompting goal-orientation as a means of coping. Arguably, however, psychological closeness and motivation to affiliate might be hard to disentangle, as psychologically close models are probably also targets of affiliative motivations. That said, experimental designs that tease apart these two factors may open up exciting new avenues for research. Since the ritual stance has important implications for group cohesion and boundary marking, paradigms in which closeness and group membership are manipulated may prove particularly insightful as a way of probing the questions set out by Kalkstein & Trope. BST would predict that, under conditions of social concern, a learner would copy irresolvably causally opaque actions more accurately from a distant in-group model as opposed to a close out-group model. Investigating these questions empirically could help to disentangle the underlying motivations and cognitive constraints that regulate copying fidelity in cultural transmission.

3.2. The role of culture

Fong, Nielsen & Legare, as well as **Clegg, Wen & Rawlings**, focused on discussing the role of culture in shaping the stances during development. The research they review suggests that the tendency to engage in one stance over the other varies across cultures. We agree that “culture is an optometrist” (Clegg et al.) and that cultural scaffolds and teaching norms encourage young learners to adopt one stance or the other. However, cross-cultural research is often correlational in nature—there are more differences between cultural groups than for example the degree to which they resort to observational learning. In order to test the impact of cultural teaching differences more directly, Western participants in an experiment may be exposed to observational learning methods for extended periods of time, while Ni-Vanuatu learners may be assigned a condition in which they receive direct instructions. Will this manipulation reverse the observed preferences or are there more cultural factors that will keep the current lens prescriptions intact? Given that the impact of culture on copying

fidelity is in line with the “cognitive gadget” account that we propose, teasing apart which cultural variables mediate stance adoption is among the most pressing questions of the BST framework. More generally, these commentaries draw attention to the bi-directional nature of the mind-culture interaction, and it is important to recognize that minds not only give rise to culture via their abilities to process, store and transmit, but that culture itself also shapes minds and the way they interact with cultural representations.

Similarly, **Puttre & Corriveau** draw attention to the role of within-culture variability, specifically the distinction between minority and majority groups in BST. They review research showing how recent immigration, being part of a religious minority as well as differing degrees of familial authoritarianism can modulate stance adoption and thus copying fidelity. Further, they propose that cultural identity may interact with developmental milestones, giving rise to different stance-propensities. These points constitute valuable additions to our framework and we wholeheartedly agree with their suggestion that future BST research needs to consider the interaction between these factors. On questions relating to social identity, we further note that BST might be particularly useful in investigating stance preferences on an even finer-grained level, namely that of individual differences. Are personality traits predictive of a social learner’s propensity to adopt one stance over the other? For instance, as discussed by **Samore and Fessler**, an individual’s tendency towards threat detection may push them towards adopting a ritual stance more readily for purposes of buffering against external risks by cleaving more closely to the group via conformism.

3.3. The costs of goal-directed action

Brown & Pain propose another set of factors that might affect stance adoption—namely specificity, riskiness and complexity of goal-directed action. They argue that the higher these properties, the smaller the margin of error, which would render copying via the ritual stance more adaptive. This is in line with many of the core assumptions of BST. For instance, we propose that the ritual or instrumental nature of an action lies in the eye of the beholder. If a sufficient amount of conventionality cues push a learner towards adopting the ritual stance when copying a food processing procedure (where mistakes are costly due to the risk of poisoning), then a high level of fidelity can be maintained even when the procedure is seen as instrumental from an outsider’s perspective (to obtain food). Adherence to local traditions via social motivations could be crucial to guarantee that particularly specific, risky or complex instrumental sequences are preserved via faithful transmission. Past

research has similarly found that ritualization in moderation may enhance the memorability of a goal-directed action (Kapitány, Kavanagh, Whitehouse, & Nielsen, 2018). But despite the applicability of BST in the domain of complex goal-directed action, it is important to emphasize that our framework does not claim that the instrumental stance results in overall low-fidelity learning. Rather, as elaborated further below, we argue that while high levels of fidelity can be maintained, an instrumental stance produces *relatively* lower-fidelity learning compared with behaviour copied via the ritual stance.

Vélez, Wu & Cushman draw attention to the cognitive costs of figuring out a model's intentions. They propose that if the costs of inferring intentions are not sufficiently offset by the knowledge that is gained as a result, then unquestioning and faithful copying becomes adaptive. Much like in Brown and Pain's commentary, this line of reasoning seeks to explain cases where instrumental action is imitated with high fidelity (where the cost of inferring model intentions is too high), while also attempting to account for low-fidelity copying within the ritual domain (whereas the intentions of the model are inferred, increasing the learner's flexibility in attaining the inferred goal by substituting steps in the sequence). As before, the interaction between social concerns and costs of inference can be investigated empirically by applying our Cultural Action Framework (Figure 3). BST would predict that ostracized learners copy an action sequence with an inferred end-goal more accurately if it is assumed to be irresolvable (the action-outcome link cannot be explained via natural causation) as opposed to resolvable (via the implicit assumption that there is a physical causal explanation for how the goal is achieved). Nonetheless, Vélez et al. would predict that an action sequence with an irresolvably opaque structure but with a salient end goal is copied less accurately than an opaque sequence without inferred end-goal, as the presence of an inferred goal might increase the willingness of learners to innovate. In short, exploring the cost-benefit tradeoffs of model inference and goal-salience can provide further nuance and insight into the cognitive processes underlying the stances.

3.4. The role of repetition

Perez proposes that the continuous performance of an action sequence may play a fundamental role in stance selection, in that it shifts the learner's attention away from the end-point rendering the steps that lead up to it more salient. We agree that repetition is likely to be a crucial factor in stance selection and advocate for its inclusion in our framework. For instance, religious practices that are part of the doctrinal mode of religion, serving as identity markers of large imagined communities (Whitehouse, 2018), often become fixated through processes of repetition, which are likely to promote the adoption

of a ritual stance and hence minimize the likelihood of deviations (Whitehouse, 2004). Further, we welcome the author's efforts to connect BST to a broader range of research in psychology and neuroscience by mapping it onto the distinction between goal-directed and habitual behaviour. We feel that the mapping is not quite as neat as Perez suggests. When stance selection is automatic, it would not involve ascription of intentions to the model, and when it is deliberative, gestural or intrinsic goals may be ascribed – for example, the intention to enact a certain sequence of body movements rather than to have a certain effect on the world. However, we regard the lessons that Perez derives from the literature on actions and habits as very valuable indeed. The temporal features of observed action, frequency of repetition, and variability of enactment may well play significant roles in stance selection.

4. Addressing the Critics

Several commentaries follow a common pattern, suggesting that some aspects of our theory are more contentious than others or were articulated insufficiently in the first place. In particular, several commentaries took the position that all cultural learning is instrumental and that BST fails to acknowledge the complexity of social learning. This has led some to conclude that a bifocal stance arrangement might be a less relevant feature of social learning than proposed in our article. We respectfully disagree with this conclusion and the arguments leading to it but we are grateful for this opportunity to clarify our position.

4.1. Culture and social learning is not always instrumental

Even though some commentaries seek to formulate accounts that are more parsimonious than BST, the simplicity of these frameworks is achieved at the cost of either ignoring certain key issues or underestimating the mechanisms involved in cultural transmission. Most commentaries that fall prey to these problems propose some variant of the assumption that all cultural practices are viewed through the instrumental stance by default. For instance, **Zentall** claims that there is no need to distinguish between instrumental and affiliative copying behaviour, because affiliative rewards are instrumental (in fact, he proposes that all types of rewards, without exception, are instrumental). He makes the argument that food, cartoons, stickers, and social praise in developmental research are all instances of the same thing, namely instrumental rewards. We find this to be, at best, a point about

semantics – an insistence that the term ‘instrumental’ should always and only be used as it is in the literature on animal learning - and, at worst, an argument which conflates very different phenomena. What we mean by instrumentality, as specified in our article, is the achievement of a goal via physical causal pathways. Instrumentality refers to the attainment of technical rewards which we argue are different from social rewards, which are less tangible and require a different degree of copying fidelity to be obtained. We assume that an adaptive organism would modulate its behaviour whenever possible based on the nature of the reward that is expected. We do not deny that the desire to affiliate is technically a goal, but we do argue that different goals (social and asocial, animate and inanimate) are reached in different ways.

Zentall offers a cognitively complex interpretation of overimitation in which the social learner is motivated to signal to the model that they are capable of accurately reproducing an observed action sequence. In fact, as we point out in our article, it is an open empirical question to what extent overimitation involves this kind of mentalising and conceptual grasp of social norms, rather than learned associations between slavish copying and social rewards. Moreover, Zentall leaves the question unaddressed as to when and why learners decide to “impress” a model via accurate copying. If instrumental learning underlies all cultural learning then how can the historical, as well as empirical (e.g. Watson-Jones et al., 2016) patterns of differential transmission be explained? It seems that, rather than offering improved conceptual tightness, Zentall’s account is based on using the term “instrumental” as a linguistic catch-all term for heterogeneous phenomena.

The critique presented by **Dubourg, Fitouchi & Baumard** also assumes that instrumental copying alone can account for conventional stability and technological innovation but does not explain how. This commentary describes examples of how sports games and social etiquette have been changed in order to satisfy non-affiliative goals, leading them to the conclusion that a bifocal theory of cultural evolution is unwarranted. There are several problems with this view. First, the examples of deliberately changing rules and etiquettes do not serve as suitable cases of social learning. Sitting on a committee responsible for changing the “off-side rule” in football to make the sport more enjoyable does not capture well the features of cultural learning environments that BST seeks to investigate.

Further, as we explore in the article, deliberate reasoning about the purpose of a rule in sport may entail a focus on instrumental goals, but these attributed purposes may significantly depart from the practice’s evolved function. An imagistic ritual, for example, may be ascribed a variety of purposes by its practitioners (e.g. appeasing the gods, turning a boy into a man, etc.), but its evolutionary function

is to create a sense of oneness among the members of localized groups (Whitehouse, 2018). Thus, practices are preserved as they are part of the cultural repertoire that marks the learner's group identity and deviating from them may achieve the same instrumental results, such as producing a soothing effect or demonstrating physical skill, but will eventually lead to weaker delineation of group membership.

Packer & Cole make a similar argument by claiming that institutional ceremonies are more than mere ritual because they are attributed a consciously formulated goal (e.g. conferring rank). Again, we disagree that any procedure with an end-goal must be instrumental. We argue that in the case of some actions that are attributed salient end goals, the causal opacity is assumed to be irresolvable and we refer to these as magical or 'quasi-instrumental' rituals, arguing that they are viewed through the lens of the ritual stance (see Whitehouse, 2011 and 2021). Accordingly, while magic relies on the presence of instrumental aims, such as warding off misfortune, it is not the same as purely instrumental action as our Cultural Action Framework makes clear. More broadly, however, Packer and Cole argue that we do not pay enough heed to the cognitive mechanisms involved in social learning and imitation. One of us has devoted decades to studying these mechanisms (e.g., Heyes, 1994; 2012, n.d./2022), but they are not a focus of the current article because stances are not reducible to social learning and imitation. Stances depend on motivational, attentional and (possibly) executive processes that differentially recruit mechanisms of social learning and imitation. This is made clear in Sections 2 and 5 of the target article.

Ze Hong advances a more measured critique in claiming that *most* magical practices are viewed through the lens of the instrumental rather than ritual stance. Firstly, we think that the question of whether *most* practices classified as 'magical' are seen as instrumental *most of the time* requires systematic quantitative investigation; it cannot be established by a few examples. Further, Ze Hong also seems to base his argument on the assumption that the presence of an end-goal (e.g. efforts to make it rain) will cause social learners, in quite deterministic fashion, to assume that the causal opacity of a practice is resolvable, prompting them to innovate and experiment with the sequence via an instrumental stance. However, as we point out in the target article, BST proposes that the resolvability of opacity also lies in the eyes of the beholder. This means that, as in the example of Sylvia's recipe in which cutting off both ends of a joint can be viewed either through an instrumental or ritual lens, an action sequence that is part of a rain making ceremony might be interpreted as either resolvable (instrumental and thus "technological") or irresolvable (ritualistic). Thus we do not strip magic of its instrumental properties as our Cultural Action Framework is sensitive to both possibilities. In fact, a

core principle of BST is the proposition that the learner's perception does not need to align with the objective reality of the action sequence that is copied. Magic can be perceived as technological and technology may be seen as magical. Accordingly, BST proposes that adopting a stance relies on a variety of factors, such as the properties of the action to be copied as well as the characteristics of the model (Figure 1 in the target article). Assuming that an end-goal always (or *mostly*) results in the adoption of an instrumental stance potentially underestimates the role and frequency of behaviour that is interpreted as irremediably opaque for purposes of generating social glue. Nonetheless, we find Ze Hong's commentary, especially the cited research about rain making practices insightful as it clearly highlights the importance of discussing both emic and etic functions of cultural practices.

A similar issue is raised by the commentary from **Fong, Nielsen & Legare** who seem to conflate arbitrariness with instrumental behaviour. It appears that from their point of view, twirling a stick before it is used to reach an object is *arbitrary* but causally transparent because there is an end-goal, and only twirling the stick without using it to reach an out-of-bound object can be seen as causally opaque. The claim that arbitrary actions with a salient end-state are causally transparent would mean that social learners always parse action sequences at the highest level, clustering all the components together and interpreting the whole sequence as either categorically opaque or transparent. This neglects cases where an action sequence with an end-state can be causally opaque in irresolvable ways, such as the case of magical practices.

Moreover, we do not follow Fong et al. in assuming an objective and sharp distinction between instrumental and ritual actions which are therefore always classified as such in deterministic fashion. Such an *action-centric* rather than *learner-centric* view potentially neglects the distinctions between emic and etic functions and leaves unaddressed the level of deliberateness by which learners come to adopt either stance (one of the focal points of BST). We feel that these distinctions mark important departures from the perspective of Fong et al. (as well as previous work by Legare and Nielsen), despite the fact that we all use similar terminology owing in part to a shared history of collaboration between our lab groups. That said, we welcome Fong et al.'s review of evidence that cultural factors can influence the development of stance behaviour. Like the research highlighted by Puttre and Corriveau, and Clegg et al, this evidence makes it plausible that stance psychology has been shaped predominantly by cultural evolution; that it is closer to the cognitive gadget than to the cognitive instinct end of the continuum.

4.2. Cued relevance and pedagogy are compatible with BST

Altinok, Tatone, Király, Heintz & Gergely propose that copying fidelity is modulated by attention to ostensive behaviours, such as eye contact and child-directed speech. Within this framework, the presence of communicated relevance encourages children to form the expectation of acquiring important knowledge, thus increasing their copying efforts. However, a model can communicate different goals, such as instrumental and normative ones. For instance, as mentioned in our article, Clegg and Legare (2016) examined the effect of goal-focused versus conventional language cues on children's imitative fidelity of a necklace-making activity, finding that the latter instructions improve accuracy of transmission over the former. Thus, the use of goal-directed and normative language as a cue is recognized by our framework (see Figure 1). It is not surprising that, all things being equal, ostensive behaviour causes heightened copying fidelity as it captures the learner's attention.

Despite the importance of ostensive cuing in social learning, we find it difficult to reconcile the natural pedagogy theory (Csibra & Gergely, 2009) with findings showing that ostracism modulates the copying fidelity with which children reproduce an action sequence (Watson-Jones et al., 2014; Watson-Jones, Whitehouse, & Legare, 2016). With communicated relevance being equal between conditions, what drives the differences in imitative fidelity if not exclusion from the ball passing activity? It is not clear to us that natural pedagogy can provide an answer to that question—nor why the two theories should be seen as mutually exclusive. The theory of natural pedagogy implies that infants use ostensive cues to make inferences about the model's communicative intentions. Thus, in BST terms, it is committed to the idea that, even in infancy, copying fidelity depends on highly deliberative, mentalistic processes. BST regards this as an outstanding empirical question. The work on “rational imitation”, cited by Altinok et al. and by Fong et al., has been challenged (e.g. Beisert et al., 2012; Heyes, 2016), but, as we indicate in section 5 of our article, claims about the deliberativeness and innateness of stance behaviour are thoroughly testable. More precisely, BST would predict that a non-ostensive condition in which learners are ostracized will produce higher copying fidelity than an inclusive condition in which ostensive cues are present.

Moreover, we find that Altinok et al.'s account is related to an important question in BST: Do social learners copy ritual actions more accurately than instrumental actions during memory formation (due to heightened levels of attention for one over the other) or are both actions encoded equally but differences emerge at the retrieval stage, during which the individual accesses the memory? Addressing this question is crucial in mapping out the cognitive structure of the bifocals and we find

that Altinok et al.'s commentary makes an important step into that direction by discussing attention-related processes such as ostensive cuing.

Nonaka, while in agreement with the basic premise that two distinct stances are adopted in social learning, disagrees that the stances are cued via the relative salience of end-goals or causal structure in the actions themselves. Albeit less fully specified, Nonaka's argument resembles that proposed by Altinok et al. Nonaka proposes that if the goal salience of food intake prompts infants to adopt an instrumental stance then there would be no reason to use spoons instead of their fingers. There are some problems with this argument, however. First, the starting assumption that spoon use is perceived as less efficient than using fingers is questionable, as it should be possible to gauge that a spoon can fit more food and will cause less spillage, hence prompting an instrumental stance. Secondly, as elaborated at length in our account, stances are not only triggered by properties of the observed actions. Their activation can also rely on a number of other cues, such as the number of models, as well as the use of instructions and normative language (see Figure 1 in the article). In this context, we see Nonaka's commentary (along with those of Altinok et al., Brown and Pain, Campbell and Fonagy, Fong et al., Lai and Stapleton, and Veluz et al.) as valuable in highlighting the many ways in which teaching, and cultural learning more generally, shapes the development of stance psychology.

4.3. Social learning is complex and copying fidelity is relative

Both **Packer & Cole** and **Buskell & Charbonneau** claim that BST draws conclusions about cultural transmission from simple imitation paradigms, which either fail to take into account the intergenerational aspect of cultural evolution or neglect complex properties, such as teaching, as a core mechanism of social learning. Despite Packer and Cole's critique, our definition is mindful of the fact that social learning constitutes a complex process which transcends simple one-to-one copying. Indeed, as Whiten points out, one of us has written at length about the need for "faithful retention" and "recurrent fidelity" (Heyes, 2018). Teaching falls well within the process of "information acquisition through interaction with others" as we envisage it and we do not agree that the complexity of social learning is underrepresented in our framework. On the contrary, BST seeks to draw attention to many different aspects of a learning interaction that can inspire the adoption of either instrumental or ritual stances, including factors such as the number of models and their characteristics (see Figure 1).

Further, the fruitfulness of our approach is evidenced by the wide range of topics and domains that BST has been applied to in the commentaries above. In fact, as observed by **Bazhydai & Karadag**, BST proves informative in not only accounting for selectivity during copying but also during teaching. Their discussion of how children's use of normative language in teaching scenarios reflects a ritual stance is compelling because it opens up a host of new questions regarding the motivations of models in learning scenarios. Do demonstrators make increasing use of conventional cues whenever group cohesion is important? Rather than oversimplifying the process of social learning, BST actively encourages its application to novel areas of research.

Buskell & Charbonneau propose that BST bases its arguments about intergenerational transmission on single instances of learning. Our framework does not assume that cultural evolution operates on dyadic one-off transmission scenarios alone, but rather draws from a variety of developmental research paradigms. Moreover, the usefulness of Buskell and Charbonneau's toy example is limited as it is based on fixed learning profiles as well as an instrumental stance which only dispenses with causally opaque elements while replicating the instrumental parts perfectly. This constitutes a coarse-grained conceptualization of transmission accuracy, which does not capture the key predictions that BST is setting out to test. We argue that adopting an instrumental stance creates *relatively* greater openness to innovation than adopting a ritual stance. As such, when approaching an action sequence through the lens of the instrumental stance one increases the likelihood of fine-grained deviations, making improvements via error more likely. Thus, Buskell & Charbonneau's hypothetical thought experiment produces the same cultural stability for instrumental and ritual actions because it was constructed with unrealistic base parameters (equipping learners with inflexible learning profiles and assuming that instrumental learning means mere elimination of causally opaque elements).

We find the issue of granularity in comparing instrumental to non-instrumental action sequences intriguing and agree that future research needs to be mindful of the level at which actions are compared. Nonetheless, it is important to note that the findings we review in our article are based on experiments that already make use of carefully controlled lab conditions (which match the level of granularity with which actions are compared). Differences in copying fidelity as prompted by ostracism threat (Watson-Jones et al., 2016), for example, cannot be explained by "the grain at which traits are learned". In light of these points, we disagree with their notion that there is an explanatory gap between learning and copying fidelity in BST.

While we cannot agree with Buskell and Charbonneau's claims, we like their concept of a "heuristic explanatory strategy". From the perspective of a cognitive psychologist, accustomed to thinking of mechanisms as "the inner machinery of agents", nearly all constructs in research on social learning and cultural evolution – from "stimulus enhancement" and "shared intentionality" to "attractors" and "cultural cognitive causal chains" – could be characterised as heuristic explanatory strategies. What is distinctive about BST is that it is explicitly designed to facilitate experimental work, of the kind described in Section 5 of our article that will elucidate "the inner machinery of agents".

Moreover, BST is sufficiently versatile to account for patterns in vertical (intergenerational) and oblique teaching, as well as horizontal transmission, such as child-led teaching, as discussed by **Bazhydai & Karadag**.

5. Conclusions and future directions

The commentaries on our target article amply demonstrate the relevance of BST to a great variety of domains, ranging across psychopathology (Campbell & Fonagy), music (Loui & Margulis), language (Scharinger & Erfurth), philosophy (Lai & Stapleton), and modelling (Leibo et al.). Collectively, the commentaries point to considerable potential for expanding our framework (Brown & Pain; Vélez, Wu & Cushman; Perez), for instance by allowing us to explore the bi-directional relationship between mind and culture (Puttre & Corriveau; Clegg, Wen & Rawlings; Fong, Nielsen & Legare), as well as to test our key arguments against those of competing accounts (Altinok et al.; Thomas, Radkani & Hung; Kalkstein & Trope). Given the capacity of BST to address the differential patterns of cultural evolution that bolster our species' success, the framework lends itself to theorizing about our evolutionary past as well as our uniqueness in the tree of life (Veit & Browning; Samore & Fessler; Whiten). Our framework is ripe with new research directions that could be explored via collaborations between authors in all the above fields. Lastly, some of the concerns raised have usefully allowed us to clarify aspects of our approach.

Funding Statement: HW's work on this response to commentaries was supported by an Advanced Grant from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement 694986).

Conflicts of Interest: None.

References

- Beisert, M., Zmyj, N., Liepelt, R., Jung, F., Prinz, W., & Daum, M. M. (2012). Rethinking “rational imitation” in 14-month old infants: A perceptual distraction approach. *PLoS ONE*, 7(3), e32563. doi:10.1371/journal.pone.0032563
- Boyer, P. (2002). *Religion explained: The human instincts that fashion gods, spirits and ancestors*. New York: Vintage.
- Clegg, J. M., & Legare, C. H. (2016). Instrumental and Conventional Interpretations of Behavior Are Associated With Distinct Outcomes in Early Childhood. *Child Development*, 87(2), 527–542. <https://doi.org/10.1111/cdev.12472>
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2009.01.005>
- Heyes, C. M. (n.d.). Imitation and culture: what gives? *Mind and Language*.
- Heyes, C. M. (1994) Social learning in animals: Categories and mechanisms. *Biological Reviews*, 69, 207-231.
- Heyes, Cecilia. (2012). What’s social about social learning? *Journal of Comparative Psychology*, 126(2), 193–202. <https://doi.org/10.1037/a0025180>
- Heyes, C. M. (2016). Born pupils? Natural pedagogy and cultural pedagogy. *Perspectives on Psychological Science*, 11, 280-295.
- Heyes, Cecilia. (2018). Enquire within: cultural evolution and cognitive science. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1743), 20170051. <https://doi.org/10.1098/rstb.2017.0051>
- Hughes, S. M., Harrison, M. A., & Gallup Jr, G. G. (2007). Sex differences in romantic kissing among college students: An evolutionary perspective. *Evolutionary Psychology*, 3(5), 147470490700500320.
- Jankowiak, W. R., Volsche, S. L., & Garcia, J. R. (2015). Is the romantic–sexual kiss a near human universal? *American Anthropologist*, 3(117), 535–539.
- Kapitány, R., Kavanagh, C., Whitehouse, H., & Nielsen, M. (2018). Examining memory for ritualized gesture in complex causal sequences. *Cognition*, 181, 46–57. <https://doi.org/10.1016/j.cognition.2018.08.005>
- Legare, C. H., Wen, N. J., Herrmann, P. A., & Whitehouse, H. (2015). Imitative flexibility and the development of cultural learning. *Cognition*, 142(512), 351–361. <https://doi.org/10.1016/j.cognition.2015.05.020>
- Nyhof, M., & Barrett, J. (2001). Spreading Non-natural Concepts: The Role of Intuitive Conceptual Structures in Memory and Transmission of Cultural Materials. *Journal of Cognition and Culture*,

1(1), 69–100. <https://doi.org/10.1163/156853701300063589>

Sosis, R., Kress, H. C., & Boster, J. S. (2007). Scars for war: evaluating alternative signaling explanations for cross-cultural variance in ritual costs. *Evolution and Human Behavior*, 28(4), 234–247. <https://doi.org/10.1016/j.evolhumbehav.2007.02.007>

Watson-Jones, Rachel, Legare, Cristine H., Whitehouse, Harvey, & Clegg, Jennifer (2014). Task-Specific Effects of Ostracism on Imitation of Social Convention in Early Childhood. *Evolution and Human Behaviour*, 35(3): 204–10.

Watson-Jones, R. E., Whitehouse, H., & Legare, C. H. (2016). In-Group Ostracism Increases High-Fidelity Imitation in Early Childhood. *Psychological Science*.
<https://doi.org/10.1177/0956797615607205>

Whitehouse, H. (2004). *Modes of Religiosity: A Cognitive Theory of Religious Transmission*. Walnut Creek: AltaMira Press.

Whitehouse, H. (2011). The Coexistence Problem in Psychology, Anthropology, and Evolutionary Theory. *Human Development*, 54: 191–9.

Whitehouse, H. (2012). Ritual, Cognition, and Evolution. In R. Sun (Ed.), *Grounding the Social Sciences in the Cognitive Sciences*, pp. . Cambridge, MA: MIT Press.

Whitehouse, Harvey. (2018). Dying for the group: Towards a general theory of extreme self-sacrifice. *Behavioral and Brain Sciences*, 41, 1–64. <https://doi.org/10.1017/S0140525X18000249>

Whitehouse, H. (2021). *The Ritual Animal: Imitation and Cohesion in the Evolution of Social Complexity*. Oxford University Press.