A Domain-General Adaptation for Tribal Value-Acquisition

Peter Carruthers
University of Maryland, College Park, USA
Email: pcarruth@umd.edu
Webpage: Peter Carruthers

Short abstract

This target-article proposes a solution to a puzzle: why is it that, across a wide range of domains, evaluative beliefs are apt to shift our evaluative experience in both short-term and long-term ways? And why are these top-down influences on affective valuation so powerful? The explanation is that it was a vitally-important adaptive problem for our hunter-gatherer ancestors to swiftly acquire the values of the tribe, including not just tastes in food, fear of local predators and dangers, and so on, but also a whole suite of local norms, as well as a default positive valuation of co-tribal members themselves.

Long abstract

This target-article proposes a solution to a puzzle: why is it that, across a wide range of domains, evaluative beliefs are apt to shift our evaluative experience in both short-term and long-term ways, and to do so powerfully? The explanation is that it was a vitally-important adaptive problem for our hunter-gatherer ancestors to swiftly acquire the shared values of the tribe. Five kinds of top-down effect are discussed. Three might have begun as domain-specific. One concerns the impact of beliefs about value – as indicated by the price of a wine or a short narrative about a novel social group – on subsequent evaluative experience. This might have started as an adaptation for acquiring the accumulated evaluative wisdom of the tribe (which things are good to eat, which nearby tribes are hostile). The second enables swift internalization of the shared norms of the tribe, leading one to value compliance with those norms for its own sake. And the third is the minimal-group effect, which immediately issues in positive evaluation of anyone seen as a member of one's own tribe (one's own in-group). In contrast, two additional top-down effects are shown to be much harder to explain as domain-specific adaptations, and are best seen as by-products of a domain-general mechanism. One is the choice effect, in which the

mere belief that one has chosen one thing over another influences one's subsequent evaluation of those things. This is proposed to result from tacit interpretation of one's own choice behavior as reflecting a difference in value. The second concerns placebo and nocebo effects on the painfulness of sensory pain. It is argued that this, too, is unlikely to be directly adaptive.

Keywords

choice effect, emotional reappraisal, minimal-group effect, norm internalization, placebo effect, value learning

Word-counts

Short abstract: 100. Long abstract: 276. Main text: 10,230. References: 2,870. Total count: 13,568.

Main Text

1. The puzzle

In later sections we will review a variety of top-down effects of belief and expectation on affective value. These effects are generally large. For instance, a meta-analysis of placebo effects on depression showed that placebos result in an improvement in depressive symptoms of 40 percent or more (Khan et al. 2012). And a meta-analysis of the effects of placebos on sensory pain found very large effect sizes (Forsberg et al. 2017). Yet one might have predicted that value-perception in general needs to be reliable. If, as many of us think, pleasure and pain (reward and punishment) are ultimately "the stars by which [all animals] steer" (Gilbert & Wilson 2007), then one might expect that it would be just as important for affective valuation to reliably track things that are genuinely valuable (that are apt to contribute to one's inclusive fitness, at least in ancestral conditions), as it is to reliably track the properties of the surrounding environment through vision, touch, and audition. For so-called "primary" (innate) rewards and punishments have been fixed by evolution to track things and activities that promote an animal's inclusive fitness; and the evaluative learning mechanisms that create secondary reward-values have likewise been selected to promote inclusive fitness.

Indeed, one of the main arguments for insisting that perceptual systems are encapsulated from central cognitive processes has always been that perception needs to reliably track the properties of the environment (Fodor 1983; Berke et al. 2022). For if perception could be influenced by one's beliefs then it would be vulnerable to all of the faults that attend fallible processes of belief-formation. And although some have claimed to find significant top-down effects of belief on perception, many of those findings have been convincingly debunked (Firestone & Scholl 2016). While it remains plausible that one's expectations have *some* effect on the contents of perception, these seem only to happen at the margins and are quite minor in nature (Ogilvie & Carruthers 2016). Why, then, should value-perception be so different? My goal in what follows is to explain why top-down influences are permitted to have such a powerful impact in the domain of value.

Before embarking on that task, however, it needs to be emphasized that the kinds of expectancy that generate the top-down effects we will be considering are entirely distinct from the predictive signals (also called "expectancies") that drive regular forms of evaluative conditioning. As we will see, the brain networks involved are largely distinct (dorsolateral and ventrolateral prefrontal cortex as the source of top-down effects, a network involving ventromedial prefrontal cortex and the ventral striatum in connection with evaluative conditioning). The functional roles of the two kinds of expectancy are quite distinct, too. Top-down expectancies create both new values and changes of value in a one-off and direct manner, prior to experience of the valued thing. The role of the "expectancies" involved in conditioned learning, in contrast, is to issue in prediction-errors when matched against one's experience of the valued thing, resulting in the stored value attached to that thing being ratcheted up or down. In the discussion that follows, "expectancy" will be used to refer to the states of belief and expectation that can issue in direct top-down changes in value, unless otherwise indicated.

I should also make clear that I will be assuming that the mechanisms underlying the effects we will be discussing are innate (not learned) and apparently uniquely human. (There is little or no evidence of their existence in other animals.) Not only do many of them emerge quite early in development, but it is quite hard to see how top-down effects of belief on affective evaluation

could be acquired through regular forms of associative conditioning. (And indeed, many of them manifest immediately, without prior experience, as we will see.) So each of the effects we discuss will either need to be explained directly as an adaptation of some sort, or it needs to be shown to be a by-product of some other adaptation. In many of the cases we will consider, neither kind of explanation has previously been attempted.

2. An evolutionary rationale

There is widespread agreement across the cognitive sciences that humans are deeply cultural creatures (Sterelny 2012; Henrich 2016; Boyd 2018). We are adapted for cultural living and cultural learning. Indeed, what best explains how humans (uniquely among primates) were able to succeed in such a wide range of ecologies and physical environments around the world (from tropical grasslands, to tropical forests, to temperate grasslands and forests, to mountains, deserts, swamps, and even the frozen arctic) is not so much our distinctive intelligence, as some have claimed (Pinker 2010) – although that was, no doubt, important – but our capacity to transmit and accumulate locally-adaptive knowledge, practices, and technologies (Boyd et al. 2011). And just as one might then predict, researchers have identified a number of cognitive biases and dispositions in contemporary humans that appear to be adaptations for cultural learning, including dispositions to admire and learn from prestigious individuals (Henrich & Gil-White 2001) and what is now called "natural pedagogy" (Csibra & Gergely 2011). Even childhood pretend play has been argued to be an adaptation for acquiring cultural skills and behaviors (Adair & Carruthers 2022).

It should be acknowledged, however, that culture in the broadest sense – that is, social learning of novel behavior – is by no means unique to humans. On the contrary, it has been observed across multiple species and taxa, including bees, birds, orcas, elephants, many primate species, and especially chimpanzees (Whitehead et al. 2019; Whiten 2021). Many creatures show some of the same social-learning biases that are found among humans, too, such as differentially copying the behavior of the majority, or copying the behavior that has the greatest perceived success or payoff. So for sure there is evolutionary continuity here. But all human cultures are, and always have been, many orders of magnitude richer than those found in other animals. While

chimpanzees, for example, have been found to have a few dozen socially-transmitted behaviors (albeit not all of them shared across populations), these are mostly fairly simple in nature, ranging from methods of grooming, to use of leaves as sponges, to nut-cracking with a stone and anvil, to termite fishing with sequential use of a stout puncturing stick and a termite-extracting frilly wand. All human cultures, in contrast, are deeply imbued throughout with socially learned behaviors of various sorts, many of them involving sophisticated multi-part tools.

Humans are not just cultural creatures, however. We are also tribal animals. For the vast majority of our evolutionary history we have lived in tribal groups within which we cooperated and sought mates, marked especially by language or dialect. And a combination of archaeological and anthropological evidence enables us to approximate what tribal life was like, at least for the last 70,000 years, and probably for a great deal longer. (Humans first evolved as a separate lineage in Africa sometime between 200,000 and 400,000 years ago.) Although there was significant variation, tribes were composed, on average, of around 1,000 adults, divided into local foraging groups of about 30 (Marlowe 2005). Composition of these local bands would have changed fairly frequently as people moved to visit with relatives or because of tensions within the group (Hill et al. 2011). Moreover, recent evidence suggests that most of the genetic changes in human populations that have happened in the last 10,000 years are related to the dietary shifts that followed the invention of farming around that time, and to changes in human immune systems driven by the resulting population expansions and the existence of settled communities and cities (Mathieson et al. 2015; Kerner et al. 2023). So most human cognitive and motivational adaptations will have emerged in a tribal-living context.

For our purposes here, the most striking fact about tribal living is inter-tribal variation combined with intra-tribal homogeneity (albeit gendered homogeneity). Tribes don't merely differ from one another in language or dialect, but also in their religious or spiritual beliefs, their ritual practices, their traditions of music and dance, modes of dress and body decoration, and more. Innovations in technology and successful foraging and food-preparation practices do tend to spread gradually among tribes that occupy similar local ecologies, perhaps through occasional trading between nearby tribes (Golovanova et al. 2021; Boyd & Richerson 2022), from kidnapping of out-of-tribe females through violence (Gat 2015), or from cultural inheritance

passed down from an ancestral group. But tribes differ widely from one another in their social practices and social norms. Internally, however, tribes are highly uniform. While there are separate norms governing the behavior of men and women, many norms will also be shared by everyone. Moreover, much of the shared culture of the tribe falls within the domain of value. People will tend to like the same foods, enjoy the same dances and music, find the same modes of dress and decoration attractive, as well as being motivated to comply with and enforce the same set of norms.

The hypothesis to be explored in the remainder of this target-article is that the evaluative homogeneity of ancestral tribal living created an adaptive pressure for a fast and reliable evaluative-learning mechanism, one that could operate without needing to rely on the sorts of conditioning processes and secondary reward-learning that we share with other animals. For people who stand out, failing to share the same tastes and practices as co-tribal members, or who breach the norms of the tribe, are likely to face ridicule and ostracization (Boehm 2001; Wiessner 2005), with serious adaptive consequences (loss of potential collaborators and mates). And indeed, even infants in the first year of life expect members of social groups to behave alike and make similar choices (Powell & Spelke 2013), and even 14-month-old infants expect agents who belong to the same group to share the same food preferences (Liberman et al. 2016.)

The upshot, I will suggest, was the emergence of powerful top-down influences of expectation and belief on affective evaluation, one that operates in essentially the same way across all evaluative domains and types of affective state. So being told, or otherwise coming to believe, that something is good or bad (or coming to believe something that is suggestive of goodness or badness, such as being a medicine or being dangerous) is apt to cause one thereafter to experience it as such. While the data that I will appeal to are mostly well-known, no one, to the best of my knowledge, has previously attempted to offer a unified explanation (or in some cases, any explanation at all). The goal of this target-article is to do just that, arguing that the hypothesis of a single domain-general adaptation can unify and explain a wide range of different findings in affective science, as well as explaining findings that would otherwise remain puzzling.

3. Effects of expectation on reward-value

There are many well-known effects of expectation on experiences of pleasure and displeasure. For example, expecting a neutral odor (or even clean air) to smell like body odor makes it seem unpleasant (de Araujo et al. 2005), expecting something to taste good makes it taste better (Grabenhorst et al. 2008), expecting a touch to feel pleasant makes it more so (McCabe et al. 2008), and expecting a wine to taste better (as indicated by its price) makes its taste more enjoyable (Plassmann et al. 2008; Fernqvist & Ekelund 2014). And these effects are not just behavioral "demand" effects, but are accompanied by changes in human reward circuitry (specifically the ventral striatum and the ventromedial prefrontal cortex; Schmidt et al. 2017). It appears that the top-down influence of expectation on reward operates across all sensory domains, at least.

Moreover, among the values that are shared within the tribe will be attitudes towards neighboring tribes, varying from cautious tolerance to murderous hostility as a result of their history of previous interactions. And as one might then expect, implicit evaluations of a novel group of people can be induced by a single short narrative, thereafter surviving many rounds of counter-conditioning (Gregg et al. 2006). Likewise, evaluative statements about a novel group of people (e.g. "Squarefaces are good, Thinfaces are bad") are sufficient to induce new evaluative attitudes in children (as measured by the implicit attitudes test), whereas associative pairing with positively or negatively valenced items does not (Charlesworth et al. 2020). So the top-down effects of belief on valuation extend well beyond the sensory domain.

From these findings alone one might conclude that there is one, or perhaps two, domain-specific mechanisms for swiftly acquiring the evaluative wisdom of one's tribe. It would have been vitally important to identify the level of threat posed by members of nearby tribes, just as it would have been important to learn about local predators and poisons (Barrett & Broesch 2012). And all hunter-gatherer tribes will have accumulated extensive evaluative knowledge related to the local ecology (what things are good to eat, what food-preparation and cooking practices are safe, which plants are poisonous, the best way to forage, and so on). There would have been strong adaptive pressure to not merely learn these things cognitively, but to translate those

beliefs into affective feelings – fear at a predator, disgust at incorrectly-prepared food, and so on. I will argue, however, that when we put these findings into a broader context, the most plausible view is that the adaptive mechanism in question is a domain-general one, operating across all evaluative domains.

4. Norm internalization

All human societies are imbued throughout with *norms* – that is, things that one must or must not do. And in the kinds of small-scale hunter-gatherer communities in which humans lived from at least 200,000 years ago until the invention of agriculture a mere 10,000 years ago, there would have been little or no internal variation in the norms that govern each tribe. They would have been accepted by nearly everyone, and people would be fully prepared to enforce them, through gossip, loss of reputation, and withdrawal of cooperation in the first instance, and ultimately through violence or exile from the community. A crucial adaptive problem for an individual growing up in such a society, then, is not only to learn what the prevailing norms *are*, but to *internalize* them – coming to value compliance with them for its own sake (both for oneself and others). For merely strategic compliance and enforcement is likely to be recognized as inauthentic, as well as leading to a greater number of failures to comply when people are faced with conflicting motivations.

The upshot is what some have called an innate "norm psychology" (Sripada & Stich 2006; Chudek & Henrich 2011; House et al. 2020). This is said to be a faculty for identifying and learning the norms of the group, for storing them and then accessing them in appropriate circumstances, and for creating intrinsic motivation to comply with them and to punish those who fail to comply with them. Some form of norm psychology likely co-evolved with the emergence of widespread cooperation in human societies, much of which takes place with unrelated individuals (Hill et al. 2011; Boyd & Richerson 2022). All extant human societies depend on such cooperation, and probably always have done. This is made possible, in part, by human norm psychology and the behavior that it supports, ensuring that cheaters and free-riders are identified and punished, and providing individuals with the intrinsic motivations to inflict such punishment, as well as to act in accordance with the group's norms in the face of

temptations to avoid doing so.

The first step in building a specific instantiation of norm psychology is to identify what the relevant norms in one's community *are*. This means forming beliefs about them. Young children are quick to learn the norms of their group, and they spontaneously enforce them on peers, protesting when rules are violated (Rakoczy & Schmidt 2013; House et al. 2020). Where does children's motivation to comply with and enforce norms come from? We know, at least, that it is unlikely to emerge from processes of conditioned secondary-reward learning; it happens too swiftly and reliably for that. A reasonable hypothesis, then, is that it results from a top-down effect of belief on affective value. Coming to believe that an act of some sort is required or forbidden swiftly imbues that action with some degree of positive or negative value respectively. These values can then be further strengthened thereafter through regular forms of reward learning, when others approve or disapprove of what one has done or failed to do.

As with the effects of expectation on reward-value discussed in Section 2, it could be the case that the causal mechanism involved in norm-internalization is domain-specific, built specifically into the structure of a norm-psychology system. But since one can operationalize actions that one *must* perform as things that it would be *bad not to do*, and actions that are *forbidden* as things that it would be *bad to do*, it might instead be the case that the mechanism is the domain-general one being proposed in this target-article. (At the very least, beliefs about norms entail beliefs about value. *Must do* entails *bad not to do* and *forbidden to do* entails *bad to do*.) Parsimony then suggests that the mechanism is, indeed, domain-general – although we should be cautious about relying on appeals to simplicity in biological and cognitive domains, where systems are generally complex and multifaceted (Carruthers 2006).

5. The minimal-group effect

Another extensively-replicated finding is that people who are randomly assigned to a small group to engage in some task together immediately form positive expectations and evaluative attitudes towards their fellow group members (Tajfel 1970; Dunham 2018). Membership of these groups can be left anonymous, or they can be marked by some arbitrary property like wearing

green T-shirts. Among other effects, people preferentially allocate resources to members of their own group, as opposed to others; and they have more positive expectations regarding the character and behavior of their in-group members.

The effect has been demonstrated in five-year-old children and younger, with children immediately showing both explicit and implicit preferences for members of their in-group, better expectations of the behavior of in-group members, and even biased encoding of positive versus negative information about in-group versus out-group (Dunham et al. 2011). The effect-sizes are moderate-to-large, and are at least half as strong as gender biases (which are quite powerful at this age), and equally as strong as racial biases (Yang et al. 2022). Indeed, even one-year-old infants will form a preference for individuals who like the same foods as them, or who like the same color mittens as them, both of which can plausibly be interpreted as cues of in-group membership (Mahajan & Wynn 2012).

Dunham (2018), in his review of the literature, details a wide range of effects that follow immediately from minimal-group membership (44 in all). These include both implicit and explicit preferences and liking for one's in-group, more positive expectations of in-group members, and more positive traits attributed to in-group members; greater empathy for pain and more overall empathy for in-group members; more favoritism in costly giving, more trust, and greater willingness to overlook in-group member transgressions; greater willingness to accept testimony from in-group members, better face memory, better recognition of emotional facial expressions, and more. In addition, Hackel et al. (2017) show that not only do people provide more resources to anonymous members of one's minimal group, but the extent to which they do so correlates with the degree of activity in the ventral striatum (the brain's main reward center) when they learn that an in-group member has received something good (from whatever source). It seems that people find it intrinsically rewarding when minimal-group members do well.

These findings make good sense in light of our history of tribal living. For membership of a cooperative group of any sort would have been a powerful cue of shared tribal membership. Hunter-gatherers would often have found themselves cooperating with strangers from the same tribe. This could happen when new members join their local traveling group, when they

themselves join a new group, or when the tribe as a whole gathers for a cooperative activity of some sort, such as building a weir or driving a herd of animals over a cliff (Hill et al. 2014; Boyd & Richerson 2022). But rarely, if ever, would people have engaged in a joint activity with members of another tribe. (Exchanging goods with members of another tribe is a form of cooperation, perhaps, but doubtfully qualifies as a joint activity.) Since people's inclusive fitness would have been strongly impacted by how successfully they integrate with fellow tribal members, supporting and being supported by them in turn, one would expect intense selection for a default positive evaluation of anyone one believes to be a member of one's in-group. And that is exactly what we find in contemporary populations.

There are two possible accounts of the mechanisms underlying the minimal-group effect, however. One is that it is yet another instance of the kinds of top-down influence of belief on affective value of the sort being proposed in the present article. It might be that recognizing someone as an in-group member causes one (innately) to believe that they are good and have good attributes, which in turn creates a positive affective valuation of them in a top-down manner. Alternatively, it could be that appraising someone as an in-group member directly causes (innately) a positive affective valuation of them, which in turn causes an expectation that they are good. I am unaware of any evidence that directly adjudicates between these two causal pathways. But we do at least know that human infants in their first year of life have expectations of in-group loyalty and support (Jin & Baillargeon 2017), suggesting that the top-down route is a possibility.

Infants as young as 5 months show a preference for someone speaking their own language over someone speaking a different one, however, as manifested by their extended looking towards the former (Kinzler et al. 2007). So it might be suggested that this early preference for in-group members favors the direct appraisal-based account. But it is unclear whether extended looking in these circumstances manifests a general group-based positive valuation of the individuals involved, or whether it instead reflects an increased opportunity for learning, and so manifests a form of curiosity. For infants should obviously target their learning at in-group same-language people rather than at outsiders. Consistent with this suggestion, Buttelmann et al. (2013) find that by 14 months infants will preferentially imitate (learning novel behaviors from) people who

speak their own language.

On the other hand, however, by around one year of age infants will prefer (choosing when offered) a puppet who expresses the same food-preference as themselves (Mahajan & Wynn 2012). This is more suggestive of innate in-group liking. But it remains unclear whether this effect is genuinely group-based in the sense of tribal group, or whether it might instead be an innate liking for any agent who is seen as sharing their *affiliative* group (where similarity-to-self would be a proxy for local care-givers). And indeed, Liberman et al. (2016) found that 14-month-old infants expect agents who affiliate with one another to share the same food preferences. So positively evaluating agents who share their food-preferences makes good sense in light of the fact that hunter-gatherer infants are routinely cared for by a network of alloparents who constitute a subset of the local group as a whole (Hrdy 2009; Chaudhary et al. 2023) – roughly speaking the mother and her friends. And we know that infants will draw inferences about affiliative relationships from a number of factors, including third-person observation of one person imitating another (Thomas et al. 2022; Kudrnova et al. 2023). So we have no unambiguous evidence of an innate mechanism that leads directly from an appraisal of own-tribe membership to positive affective evaluation.

In contrast, we do know that infants as young as 9 months have general expectations about group membership, where the individuals involved are animated agents (none of whom belong to the same group as the infant), and where group membership can be marked by language, joint action (such as dancing together), or affiliating together (Liberman et al. 2016). In particular, they expect speakers of the same language (but not speakers of different languages) to affiliate with one another in a friendly way (Liberman et al. 2017). And in a variety of experiments done with infants aged between 12 and 17 months, they expect individuals belonging to the same group to help one another (but not to help someone from a different group) and to come to the assistance of a group member who is in conflict with a member of another group (Jin & Baillargeon 2017; Ting et al. 2019; Pun et al. 2021).

Since it is unlikely that infants at these young ages have any experience of the behavior of members of different social groups (or even evidence from which they could infer that there are

such things as social groups), it seems that they possess innate expectations concerning the significance of group membership. In particular, since they expect group members to help one another and support one another when needed, it is plausible to think that identifying someone as belonging to one's own social / tribal group would lead them to believe that the person is good and helpful (as, indeed, we find in minimal-group experiments with older children and adults). And then a domain-general system that creates positive affect from positive expectations would produce immediate liking for in-group members (as we also find in minimal-group experiments).

I conclude that it is at least possible that the minimal-group effect on in-group evaluation results from a top-down influence of evaluative belief on affective value. And there is some reason to think that this is the most plausible explanation. But there is nothing here to imply that the mechanism involved is a domain-general one. One the contrary, it might have evolved specifically in the domain of tribal-membership evaluation, because of the adaptive advantages of a default positive evaluation of one's own tribal members. Since hunter-gatherers would often have found themselves engaging in cooperative activities with strangers who are nevertheless members of the same tribe, it would have been adaptive to evaluate them positively by default, issuing in greater trust, willingness to share, and so forth.

It should be noted, however, that mechanisms that initially emerged under domain-specific selection pressures (for acquiring accumulated tribal evaluative wisdom; for norm internalization; and perhaps for positive in-group-member evaluation), and that emerged at different times in human evolution, might nevertheless have resulted in a mechanism that now operates across all evaluative domains. Even if one accepts a domain-specific evolutionary history, it remains possible that the eventual product of that history was a domain-general mechanism that was constructed piecemeal. And this will become more plausible the more instances can be found of top-down effects on value that are best seen as by-products of a more general system rather than as directly adaptive. This is where we go next.

6. The choice effect

This is a well-known and massively replicated finding. People who are forced to choose between

two equally-rated options thereafter shift their evaluations of those options accordingly – liking the chosen item more strongly and/or liking the rejected item less (Brehm 1956; Enisman et al. 2021). And just as with the expectation-based effects reviewed above, these effects aren't merely behavioral in nature, but also involve changes in underlying neural evaluation networks (specifically in the ventral striatum; Sharot et al. 2009). They also give rise to long-term evaluative change, even being discernable three years later (Sharot et al. 2012). A range of domain-specific explanations for the choice effect have been offered over the decades, including post-choice motivated reasoning, biased reflection on the properties of the two options, or as resulting from an attempt to protect one's sense of self-integrity or self-esteem. But these explanations seem to be ruled out by a number of more-recent findings.

The choice effect continues to operate in people with severe amnesia (who won't remember their earlier choices for more than a minute or two) and in people who are placed under cognitive loads that prevent reflection (Lieberman et al. 2001). The effect can also be found in preschool children, who are unlikely to engage in post-choice reasoning or reflection (Egan et al. 2007). Moreover, the effect is just as powerful when people are tricked into *believing* that they have chosen one thing over another (without really having done so) or when they choose blindly, in ignorance of the identity of the two options (Egan et al. 2010; Sharot et al. 2010). So the effect cannot result from undetected pre-choice differences in evaluation, as confirmed by Enisman et al. (2021). Even more striking, the effect of believing one has made a choice on subsequent evaluation is found among human infants in their second year of life (Silver et al. 2020), and the effect of blind choosing has been established early in the third year, and turns out to be unrelated to capacities for self-identification (Wiesmann et al. 2022).

Given these findings, I suggest that the best (albeit currently untested) explanation of the choice effect is as follows. One's mentalizing system takes as input one's own observed behavior (as it always does; Carruthers 2011), and deploys something like the principle, *if S chooses A over B then A is probably better than B*, or perhaps, *if S chooses A over B then S prefers A to B*. The latter is known in economics as the principle of "revealed preference," and plausibly plays an important role in observational value-learning – observing another agent choose one thing over another, one infers that the chosen option is likely to be the better or more desirable of the two.

So when an adult, child, or infant chooses one thing over another for themselves they form the belief that the former is better than the latter from tacit interpretation of their own behavior, and that belief then exerts a top-down influence on the agent's subsequent evaluation of the two options. We know that infants can engage in this sort of simple mentalizing (Baillargeon et al. 2016), and we know that the mentalizing system continues to operate in the first person (Carruthers 2011), so this explanation makes good theoretical sense.

On this account, then, coming to believe that one has chosen A over B leads one to believe that A is *better* than B, and this in turn causes a change in one's relative affective valuation of A and B. It is hard to see how this could be anything other than a by-product of a more-general influence of beliefs about value on affective value. For what, otherwise, could possibly be the adaptive advantage of the choice effect? (Note that it is unlikely to be learned or result from affective conditioning, given its early emergence in infancy; nor is it at all clear how it could be learned at all.) How would it improve one's inclusive fitness if, having arbitrarily made a choice among two equally valued options (or merely believing that one has made such a choice), one thereafter values one of the options more highly than the other? For this will likely fail to reflect any real difference in value.

It might be replied that the effect evolved for its capacity to protect a positive self-conception, even if it now operates (as a side-effect) in infants who lack such a conception. But it remains unclear why a hard-to-make choice of some sort should impact one's sense of self-worth in the first place, nor why this should give rise to any significant adaptive pressure. So I conclude that the choice effect is best explained as a by-product of a more-general impact of expectations of value on affective valuation.

The choice-effect has been also found in Capuchin monkeys, however (Egan et al. 2007, 2010). Assuming that the data are robust, this might seem to raise a problem for the hypothesis being pursued in this target-article. For the choice-effect cannot, in monkeys, be a side-effect of a top-down adaption for swift acquisition of *cultural* values. In contrast with the human case, however, we have no evidence of a choice-induced *evaluative* shift in monkeys, let alone a lasting one. The monkeys may merely be self-applying a foraging-related behavior-rule, along the lines of,

"If option B has been rejected in favor of A, pick A rather than B." Something like this rule is needed to explain observational learning of foraging choices anyway, which we know monkeys are capable of.

Moreover, it is quite unclear why there would have been any adaptive pressure for the evolution of a mechanism that would take an observer directly from the observation that another monkey prefers A to B to immediate updating of the value that the observer attaches to the two options. For in almost all real-world situations the observing monkey can readily sample the chosen option, respond accordingly, and start acquiring a new conditioned value. It seems that a foraging-related behavior-rule is all that the monkeys really need. In the absence of evidence of a direct impact on evaluative learning, then, it is reasonable to believe that a behavior-rule is all that they really have.

One other alternative explanation should be considered. This would reduce the choice effect to the IKEA effect (Norton et al. 2012). Effortful tasks boost the value of the rewards that result. This is an effect that is found across multiple species, and is arguably best explained as a sort of contrast-effect (Alessandri et al. 2008; Johnson & Gallager 2010; Inzlicht et al. 2018). The contrast between the negative valence involved in an effortful action and the positive valence received immediately afterwards boosts the extent of the latter, issuing in a larger experience of reward. So it might be suggested that the cognitive effort involved in making a difficult choice between two equally valued items boosts by contrast the value one attaches to the option chosen.

This explanation is unlikely to be correct, however, since the choice effect is found in situations of blind choice, in which the chooser only learns after the fact which of two equally-valued items they have selected (Egan et al. 2010; Wiesmann et al. 2022). For example, participants who were told that they would be making a "subliminal" choice between two potential vacation destinations (actually just two brief flashes on the screen), who were then told that they preferred Paris to London (say), thereafter valued the former as a vacation destination more highly (Sharot et al. 2010). Since no effort at all would have been involved in the initial button press following the flashes of light ("just go with your gut," they were told), there is unlikely to have been any negative valence involved to set up a contrast.

We can conclude that the choice effect on affective value is unlikely to be an adaptation in its own right, nor does it result from some other domain-general mechanism. Rather, it is best seen as a by-product of a domain-general influence of expectations on affect.

7. Placebo and nocebo effects

The existence of placebo effects on various forms of illness has been known about for centuries. (Thomas Jefferson's doctor is reported to have said that he had prescribed more sugar-pills than real medicines over the course of his career.) And such effects can be remarkably powerful – by some estimates accounting for around 50 percent of the benefit provided by many established medications and procedures. Placebos are now known to have their strongest effects on affective forms of illness, including pain (both acute and chronic), irritable-bowel syndrome, depression, and anxiety disorders (Ashar et al. 2017; Petrie & Rief 2019). Indeed, it appears that the benefits they have for other kinds of illness may come via their positive impact on stress and anxiety, thus reducing inflammation and speeding healing (Liu et al. 2017). In fact, meta-analyses of existing studies suggest that as much as 80 percent of the benefit provided by either drug-treatments or psychotherapy for affective illnesses like depression and anxiety is accounted for by a combination of placebo effect and spontaneous recovery (Cuijpers et al. 2012; Khan et al. 2012).

There are two basic kinds of placebo effect, however (Ashar et al. 2017). One works by creating an *expectation* (or partial expectation) of improvement. (Since clinical trials with a placebo control require participants to be informed that there is only a 50 percent chance that they will be assigned to the test-medicine condition, participants should not believe outright that they are taking a real medicine; although no doubt many of them do – which may be why the strength of placebo effects correlates with trait optimism; Kern et al. 2020.) These expectations have direct and powerful effects, not just on people's reports of their affective state, but also on the neural networks that underlie reward and punishment (Ashar et al. 2017). So placebos provide an instance of powerful top-down effects of belief (or partial belief) on affective experience – as do nocebos, where expectations of a bad outcome induce one to feel worse.

The second form of placebo effect results from affective conditioning. For example, if pain is initially treated with an analgesic whose dosage is gradually reduced without the patient's knowledge, the degree of pain-reduction can be sustained once the patient is merely taking an inert pill; and the benefit remains even when the patient is informed that they are no longer receiving a real medicine (Schafer et al. 2015). Conditioned-placebo effects have been shown to be effective in non-human animals (Guo et al. 2011; Meeuwis et al. 2020), whereas there is no evidence of expectation-induced effects among animals.

Some researchers in the field have attempted to unify the two kinds of placebo effect, arguing that both are based on *expectancies* (Colagiuri et al. 2015). For evaluative learning via the conditioning route is thought to depend on evaluative prediction-errors that drive evaluative change. This would be unobjectionable if it were intended merely as a terminological convenience. But if it is meant as a substantive claim then it is surely mistaken. For as we noted in Section 1, there is a crucial functional difference between the kinds of expectancies that are embedded in the bottom-up evaluative-learning networks that we share with other animals (involving interactions between the ventral striatum and ventromedial and orbital-frontal cortices) and the sorts of expectancy that result in top-down evaluative modulation that is driven by the dorsolateral and ventrolateral prefrontal cortices, as often happens in humans. The two kinds of placebo effect are cognitively and neurally distinct, even though they can both be described using the term "expectancy."

The two kinds of placebo effect do often co-occur, however. An initial top-down-caused improvement in symptoms provides a conditioning reward-signal that enables the bottom-up system to further strengthen the effect. And conversely, the success of previous treatments can amplify the placebo effect, whereas prior failures suppress it (Kessner et al. 2013; Zunhammer et al. 2017). Indeed, it can often be unclear without separate testing whether some of the factors that can induce placebo effects in any given case are belief-based or conditioning-based, such as the benefit that can be provided by a clinician's empathic manner or perceived competence.

The strength of placebo and nocebo effects is puzzling, especially when seen through the lens of evolution. In connection with depressive and anxiety syndromes t is easy to see why signals of

social support (whether from a trusted clinician prescribing what is believed to be a medicine or from a sympathetic counsellor) might have an impact. For support from tribal members should mean that the world is a more promising and less dangerous place than one had previously assumed, shifting one's appraisals accordingly. But why should such signals have any impact on the painfulness of an arthritic knee or a twisted ankle? If the strength and vividness of the sensory component of pain is even an approximately reliable correlate of the extent of the damage or risk of damage represented, then one would think that it would be most adaptive for one's appraisal of the badness of the signal to track its intensity. If one appraises a signal of severe damage as less bad than it actually is, then one might think that one's resulting behavior could result in yet more damage – for example, if one walks unsupported on one's twisted ankle too soon. And equally, if one appraises that same signal as worse than it really is (as in the nocebo effect), one may miss out on adaptive opportunities of various sorts. In short, why should a signal of social support, or its lack, lead people to appraise their pain sensations as being better or worse than they really are?

Remarkably, very few people have taken up the challenge of explaining how placebo analgesia might be adaptive. Almost all the work has focused, instead, on the modulators of the effect and its underlying neural mechanisms (Ashar et al. 2017; Petrie & Rief 2019). One exception is Trimmer et al. (2013), who provide a model of when and why expectations should modify the activity of the immune system, ramping it up or down. (This is known to be one of the ways in which placebos can impact non-affective forms of illness.) If an individual is short of resources (e.g. facing starvation), or is in an otherwise-threatening situation, then it may be adaptive to avoid investing in a strong immune response until the threat has been dealt with. For the immune system is energetically expensive to activate and run. So conversely, anything that provides the individual with cues indicating that resources are in the offing or that threats have been reduced should enable a ramping up of the immune response, with subsequent faster healing. This explanation only works if such cues are generally reliable, however. Yet it is quite unclear why mere sympathy or concern expressed by community members should mean that it is safe to ramp up one's immune response.

Moreover, Trimmer et al. (2013) do not attempt to explain why the proposed impacts on the

immune system would have to proceed via changes in the individual's affective state, in any case. Why should signals of social support cause one to appraise one's pain experience as being less bad, with down-stream effects on the immune system? Why is the impact on the immune response not a direct one? For we know that modulation of the immune system can result from bottom-up conditioning alone (Goebel et al. 2002; Schedlowski & Pacheco-López 2010). And indeed, there is good reason to think that placebo-based immune and hormonal responses can *only* be produced via the conditioning route (Benedetti et al. 2003; Wendt et al. 2013). Mere beliefs that patients are receiving a medicine have no effect. So Trimmer and colleagues have no explanation for why beliefs should impact the affective component of pain.

Steinkopf (2015) offers a somewhat different account of why the placebo effect might be adaptive. He claims that many of the symptoms of conditions that respond to placebos such as fever, apathy, swelling, or obvious signs of pain have evolved as *signals* of the need to receive social support. Once there is evidence that such support is forthcoming, those signals have done their work and can fade. But Steinkopf, too, ties the evolution of the placebo effect to optimal functioning of the immune system. He claims that the syndromes that can be impacted most by placebos are those in which the immune system can do the most to contribute to recovery. So this account faces all the same weaknesses as the previous one.

Moreover, it is quite unclear why the negatively valenced component of affect is needed for the signaling function. All affective states produce spontaneous and directly-caused motor output that can serve as reliable signals to other people. In the case of pain, for example, it includes the pain-grimace, a tendency to groan and cry out, a tendency to nurse the relevant body component, and so on. And these behaviors are mostly served by the lateral sensory-network component of pain rather than the medial affective-evaluative one. So one would think that it should merely be the spontaneous behavioral expression of pain that would subside given signals of social support. It remains unexplained why the sensation of pain should feel less painful (less bad) as well.

I suggest, in contrast, that placebo analgesia is best seen as a by-product of a domain-general system for creating top-down effects on affective valuation generally, rather than as adaptive in its own right. Placebo effects are found for a variety of conditions, of course, including pain,

depression, anxiety, insomnia, irritable bowel syndrome, and nausea (Colagiuri et al. 2015). And although there are commonalities among the brain networks involved, there are also significant differences (Frisaldi et al. 2020). All top-down forms of placebo effect seem to involve activity in dorsolateral and ventrolateral prefrontal cortex (which are thought to code for and signal one's expectations of improvement), dorsal anterior cingulate cortex (which is an important way-station in the processing of negatively valenced stimuli), and the ventral striatum (which is heavily involved in evaluative processing generally). But pain analgesia is mediated via projections to the periductal grey (a subcortical region that projects downwards into the spinal cord) as well as the insula cortex (which is heavily involved in the processing of negatively valenced somatosensory stimuli). Placebo reduction of anxiety, in contrast, involves projections from the core network to basolateral and ventrolateral amygdala (Benedetti 2014).

Importantly for our purposes, however, Atlas et al. (2010) conducted an extensive mediation-analysis of the entire pain network. They found that the effects of expectations on all the other components of the pain network (the insula, anterior cingulate cortex, amygdala, and periductal grey) are mediated by interactions between the ventral striatum and ventromedial prefrontal cortex. This is the core system involved in affective valuation and affective learning generally, as we will see in Section 8. So the top-down impact of expectations on affective valuation may utilize that same shared network that is also involved in all other forms of top-down influence, consistent with a domain-general account.

The fly in the ointment here, however, is that while almost all of the conditions where powerful placebo effects can be found are deeply affective in nature, there is one exception: Parkinson's, which is a motor disease. However, the benefit that placebos provide in Parkinson's may be a by-product of shared dopamine projections from prefrontal cortex to the striatum. Parkinson's is often treated by boosting the patient's tonic dopamine levels, which can also have unfortunate side-effects on evaluative learning, making the patient vulnerable to addictions of various sorts (Dagher & Robbins 2009). Moreover, it turns out that placebo effects on Parkinson's disease are only seen in people whose disease had previously been ameliorated by drug treatments (Benedetti et al. 2016; Frisaldi et al. 2017). This suggests that the mechanisms are quite different. They involve the conditioned-learning route, rather than being a top-down effect of belief or

expectation.

8. Domain-general value networks

We have reviewed five different domains in which expectations of value have significant top-down effects on affective valuation. One concerns the acquisition of reward-values generally (sensory experiences, foods, groups of people, and more). The second is norm-internalization, where beliefs about the badness of actions or omissions may produce some degree of matching intrinsic valuation. The third is the minimal-group effect, which is plausibly a product of innately-caused beliefs about the default goodness of one's believed in-group. The fourth is the effect of choice on subsequent valuation, where the best explanation of the finding is that it is a by-product of a more-general impact of expectation on affective value. And then finally, there are placebo and nocebo effects on affective experience, and specifically on the painfulness of pain. The best explanation of these effects, too, is that they are a by-product of a domain-general system for top-down modulation and learning of affective values.

Taking these five sets of findings together, I submit that the best explanation of them as a whole is that the need for swift and reliable tribal value-acquisition produced selection-pressure for a domain-general system that allows beliefs about value to impact the experience of value, modulating the core evaluative network linking the ventral striatum and ventromedial prefrontal cortex, in particular. (Either that, or perhaps that a number of distinct pressures might have led initially to top-down systems operating in the domains of tribal object-valuation and activity-valuation, norm internalization, and cues of tribal group membership that gradually merged into a domain-general one.) Otherwise we lack anything for the choice effect discussed in Section 6 and the placebo and nocebo effects on pain discussed in Section 7 to be by-products *of*.

Indeed, consistent with the domain-general value-learning hypothesis being proposed here, Plasmann & Wager (2014) argue that it the very same core system that is modulated by expectations of the value and enjoyment of consumer products that also underlies placebo analgesia and other forms of placebo effect. Specifically, they argue that the impact of expectations in both domains operates via an effect on the same core evaluative network linking

the ventromedial prefrontal cortex with the ventral striatum.

Indeed, many have argued that there are closely overlapping brain networks underlying the core processing of both positive and negative affective states (Ellingsen et al. 2015; Becker et al. 2019). Even the subjective (negative) value of cognitive effort appears to be processed in the same domain-general network (Westbrook et al. 2019). And in the case of positive values, there is evidence of a common neural code for both information-value and other forms of reward (Charpentier et al. 2018). In fact, Kobayashi & Hsu (2019) show that the subjective value of both information and other forms of reward is represented in voxel-wise BOLD signals in ventromedial prefrontal cortex and the ventral striatum. Importantly, they show in addition, via cross-categorical decoding, that both sets of values share a common coding scheme.

The fact that there is a shared core system for both positive and negative valuation characterized at the network-level fails to demonstrate that there aren't distinct micro-circuits for processing distinct forms of value, of course. So it doesn't *follow* that once evaluative expectations encoded in dorsolateral and ventrolateral prefrontal cortex had acquired the capacity to directly create or modulate value in one domain (one micro-circuit in this network), that it should automatically extend to all, issuing in a domain-general mechanism for top-down evaluative learning. But it is easy to imagine how such a thing might happen; especially given the existence of individual neurons in the striatum that are involved in the representation of distinct forms of value (Bromberg-Martin & Hikosaka 2009).

I conclude that the domain-general adaptation being proposed in this target article is not just consistent with our best knowledge of the neuroscience of valuation, but is at least tentatively supported by that knowledge.

9. Alternative explanations

I have been arguing that a range of top-down influences of expectation on affect are best explained in terms of a single domain-general meta-affective adaption for cultural (originally tribal) value-learning. One alternative proposal could be that these top-down effects are not an

adaptation at all, but rather a by-product of the expansion of the neocortex in humans, and specifically the further relative expansion of the prefrontal cortex. (The latter is, as we have seen, the seat of the top-down signals that produce the affective changes in question.) If this were the case, however, then one might expect that individual differences in intelligence (fluid IQ) or executive control abilities would correlate with the strength of placebo effects. I am aware of no evidence that supports either of these possibilities. Indeed, although the literature is somewhat tangled, it seems that the strength of people's placebo responding in a variety of domains either correlates or anti-correlates with a range of personality variables, including optimism, need for cognition (thoughtfulness), pleasure seeking, and (negatively) somatosensory bodily awareness (Geers et al. 2007; Plassmann & Weber 2015; Kern et al. 2020). So there is no reason to think that the top-down effects of expectation on value are a mere by-product of cortical expansion.

It might be claimed, however, that there is an alternative unifying account available. This is that the selection pressure underlying this domain-general system was for better emotional self-management instead. Humans needed to become capable of modifying their emotions at will in order to be successful at group living, it might be said. And by intentionally intervening to improve their moods and outlook on life, they thereby reap physical fitness benefits as well. For there is substantial evidence that optimistic people are healthier and live longer (Diener & Chan 2011; Lee et al. 2019), and are more resilient in the face of adversity (Kleiman et al. 2017; Bonanno 2021). So perhaps what evolved initially was a capacity for intentionally-produced high-level conceptual representations to alter one's affective states and shift one's stored values. The effects of externally-caused expectations on affect can then emerge as a bi-product of our capacity for emotional self-management.

The central mechanism in emotional management (in addition to situation selection and attentional redirection) is *reappraisal* (Gross 2015). Initially angered by an insulting comment, for example, one reminds oneself that the person is tired and stressed after a long day, and so probably did not intend the insult, and one's anger then subsides. Potentially violent conflicts are thereby avoided. By re-conceptualizing the insult as a slip or mistake of some sort one deflates its emotional impact. This is a top-down impact of concepts and beliefs on affective experience, and one that appears capable of operating across all kinds of affective state. Thus by re-

appraising the initially-attractive smell of cigarette smoke as dirty and disgusting one can thereby diminish one's desire to smoke, for example. And so on.

To begin evaluating this idea, we need to draw a distinction between two kinds of emotional re-appraisal. One is automatic / spontaneous, and the other is intentionally controlled and directed. In the real world, reappraisal is often just a matter of looking closer, waiting briefly until the target becomes clearer, and so on. Experiencing a burst of fear at the wriggling shape in the grass near one's foot, one looks closer only to realize that it is a strand of paper that is gyrating in the breeze, or one recognizes it as a harmless grass-snake. As a result, one's fear subsides. Likewise, a gazelle that notices a cheetah walking nearby may experience a burst of initial anxiety, but pauses briefly to watch. It soon becomes apparent that the cheetah is not in hunting mode but is transporting one of her cubs to the safety of a nearby tree, or is merely passing through in the direction of the local water hole. As a result, the gazelle's anxiety subsides, and it returns to feeding. This is a familiar form of re-appraisal, but its influence on affect is not really a top-down one. The inputs to the affective system change with greater attention or alter with the passage of time. The mechanism through which it operates is entirely bottom-up, even if dependent on a decision to look closer or to watch for a little longer.

Humans, however, can intentionally modulate their affective state be reconceptualizing the object of the emotion in different terms or through intentional self-talk ("he didn't really mean it"). This is one of the main strategies underlying emotional intelligence and emotional self-management, and it *is* genuinely a top-down effect of cognition on affective experience. But in other respects is quite different from those considered in previous sections of this target-article. For those rely on one's antecedent expectations and beliefs, and are not actively undertaken or intended. They occur spontaneously, and do not depend on the agent's decision-making. So re-appraisal is not itself an instance of the general adaptation we have been considering and arguing for – the top-down influence of expectation on value. But it may nevertheless be claimed to give rise to an alternative explanation of the same range of phenomena as a by-product.

One reason to doubt this proposal, however, is that most people are not very good at emotional regulation, and need to be coached in cognitive re-appraisal techniques – through cognitive-

behavioral therapy, for example, or through learning such techniques as a strategy in a particular emotional domain (such as anger management). This is because in general one should expect systems that have been adapted for a particular purpose to be good at what they do. And one should expect, too, that an innate system should operate spontaneously, rather than being intentionally directed. Moreover, we have no evidence of the use of emotional re-appraisal in infancy and childhood. Indeed, quite the contrary, young children are notorious for not being capable of modulating their emotional reactions to things.

Another reason for rejecting the proposed explanation of top-down effects on affective value is that emotional reappraisal seems incapable of creating genuinely novel values. Rather, it works by reconceptualizing an event or stimulus in terms of something that is already valued, or by using terms that are already value-laden. Top-down imagining or re-appraisal shifts one's evaluation of the current stimulus; but the effect is ephemeral, and doesn't shift one's underlying stored valuation of it. This is why therapies that use reappraisal techniques take time to be effective – using conditioned learning, in effect. The top-down effects of belief and expectation, in contrast, immediately alter the underlying valuation of the thing or event in question. Expecting something to be good or bad can create some degree of congruent experience of it out of nothing, with subsequent effects on one's long-term valuation.

A final reason to prefer the tribal-learning hypothesis is that it is better positioned to answer the puzzle with which we began: how can top-down effects of belief on affective value be adaptive? The tribal-learning hypothesis can rise to this challenge. In part this is because many of the values that one needs to acquire in a culture are actually independently valuable, quite apart from the culture itself. Many cultural values represent the tribe's accumulated wisdom about the local ecology (what things are good to eat, what food-preparation and cooking practices are safe, which animals are dangerous, the best way to forage, and so on). Moreover, it is vital for one's inclusive fitness that one should internalize the norms that are prevalent in one's group, and that one should have a default positive evaluation of anyone whom one sees as a member of one's own tribe. These benefits may significantly outweigh any costs that attend the choice-effect (which are likely small) and placebo and nocebo effects (which may be more significant).

In contrast, telling oneself that the future will be good when really it will not be seems obviously problematic (one may not prepare adequately, for example), even if it has good effects on one's mood and optimistic outlook. (It is true that optimism seems adaptive other things being equal, as we noted earlier, but other things are often *not* equal, and a more realistic take on the future would be better). And telling oneself that an insult was not really intended when actually it was may lead to a loss of reputation, especially in small-scale societies, since it shows one can be pushed around by others without cost. And so on: emotional reappraisal can have both benefits and down-sides, depending on context (Troy et al. 2013).

10. Conclusion

I have argued that the best explanation of a variety of top-down effects of belief and expectation on affective value is that there is a single domain-general adaptation for tribal value acquisition. This hypothesis is consistent with the idea that a number of initially-distinct mechanisms may have emerged at different times in more-specific domains, perhaps beginning with mechanisms for acquiring the evaluative wisdom of the tribe and internalizing its norms, as well as a mechanism for default positive evaluation of members of one's own tribe over others. But if we accept that these separate evolutionary forces converged to produce a single top-down mechanism we can also explain both the choice effect and why placebo effects on pain should be so powerful, despite the fact that they seem not to be directly adaptive – they can be by-products of the domain-general mechanism in question.

The argument has admittedly been speculative. But to the best of my knowledge it addresses questions that have not previously been confronted. And it presents a challenge to empiricists who might want to insist that the top-down effects we have been considering are some sort of cognitive/motivational *gadget*, acquired through evaluative learning rather than being innately channeled. (How could that possibly happen? How could one construct a novel and wholly general mechanism for acquiring novel values via gradual associative conditioning?) Admittedly, too, it is hard to directly demonstrate the truth of a negative (that the choice effect and placebo effects on pain are *not* adaptive). But there is a challenge here, as well, for anyone who wants to reject the domain-general hypothesis I have proposed; namely, to demonstrate and offer evidence

of their adaptiveness. Until either of those things is done, I submit that the most plausible hypothesis is that there is a uniquely-human domain-general mechanism that enables beliefs about value to impact one's affective valuation of things, that evolved under pressure to rapidly acquire the values of one's tribe.

Acknowledgements

I have benefitted from discussion of some of this material from many of my graduate students over the years. I am grateful to Joe Gurrola for his comments on an earlier draft.

Funding & Conflicts of Interest

None.

References

Adair, H. & Carruthers, P. (2022). Pretend play: More imitative than imaginative. *Mind & Language*, 38, 464–479.

Alessandri, J., Darcheville, J., & Zentall, T. (2008). Cognitive dissonance in children: Justification of effort or contrast? *Psychonomic Bulletin & Review*, 15, 673–677.

Ashar, Y., Chang, L.J., & Wager, T. (2017). Brain mechanisms of the placebo effect: An affective appraisal account. *Annual Review of Clinical Psychology*, 13, 73–98.

Atlas, L., Bolger, N., Lindquist, M., & Wager, T. (2010). Brain mediators of predictive cues on perceived pain. *Journal of Neuroscience*, 30, 12964–12977.

Baillargeon, R., Scott, R., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, 67, 110–118.

Barrett, H.C. & Broesch, J. (2012). Prepared social learning about dangerous animals in children. *Evolution and Human Behavior*, 33, 499–508.

Becker, S., Bräscher, A., Bannister, S., Bensafi, M., Calma-Birling, D., Chan, R.C., Eerola, T., Ellingsen, D., Ferdenzi, C., Hanson, J., Joffily, M., Lidhar, N., Lowe, L., Martin, L., Musser, E., Noll-Hussong, M., Olino, T., Lobo, R.P., & Wang, Y. (2019). The role of hedonics in the Human Affectome. *Neuroscience & Biobehavioral Reviews*, 102, 221–241.

Benedetti, F. (2014). Placebo effects: From the neurobiological paradigm to translational

implications. *Neuron*, 84, 623–638.

Benedetti, F., Frisaldi, E., Carlino, E., Giudetti, L., Pampallona, A., Zibetti, M., Lanotte, M., & Lopiano, L. (2016). Teaching neurons to respond to placebos. *Journal of Physiology*, 19, 5647–5660.

Benedetti, F., Pollo, A., Lopiano, L., Lanotte, M., Vighetti, S., & Rainero, I. (2003). Conscious expectation and unconscious conditioning in analgesic, motor, and hormonal placebo/nocebo responses. *Journal of Neuroscience*, *23*(10), 4315-4323.

Berke, M., Walter-Terrill, R., Jara-Ettiger, J., & Scholl, B. (2022). Flexible goals require that inflexible perceptual systems produce veridical representations: Implications for realism as revealed by evolutionary simulations. *Cognitive Science*, 46, e13195.

Boehm, C. (2001). *Hierarchy in the Forest*. Harvard University Press.

Bonanno, G. (2021). *The End of Trauma*. Basic books.

Boyd, R. (2018). *A Different Kind of Animal: How Culture Transformed our Species*. Princeton University Press.

Boyd, R., & Richerson, P. (2022). Large-scale cooperation in small-scale foraging societies. *Evolutionary Anthropology: Issues, News, and Reviews*, 31, 175–198.

Boyd, R., Richerson, P., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, 108, 10918–10925.

Brehm, J. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, 52, #384.

Bromberg-Martin, E. & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, 63, 119–126.

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, 84, 422–428.

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press.

Carruthers, P. (2011). *The Opacity of Mind*. Oxford University Press.

Charlesworth, T., Kurdi, B., & Banaji, M. (2020). Children's implicit attitude acquisition: Evaluative statements succeed, repeated pairings fail. *Developmental Science*, 23, e12911.

Charpentier, C.., Bromberg-Martin, E., & Sharot, T. (2018). Valuation of knowledge and

ignorance in mesolimbic reward circuitry. *Proceedings of the National Academy of Sciences*, 115, E7255–E7264.

Chaudhary, N., Salali, G., & Swanepoel, A. (2024). Sensitive responsiveness and multiple caregiving networks among Mbendjele BaYaka hunter-gatherers: Potential implications for psychological development and well-being. *Developmental Psychology*, 60, 422–441.

Chudek, M. & Henrich, J. (2011). Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences*, 15, 218–226.

Colagiuri, B., Shenk, L., Kessler, M., Dorsey, S., & Colloca, L. (2015). The placebo effect: From concepts to genes. *Neuroscience*, 307, 171–190.

Csibra, G. & Gergely, G. (2011). Natural pedagogy as an evolutionary adaptation. *Philosophical Transactions of the Royal Society B*, 366, 1149–1157.

Cuijpers, P., Driessen, E., Hollon, S., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: A meta-analysis. *Clinical Psychology Review*, 32, 280–291.

Dagher, A. & Robbins, T. (2009). Personality, addiction, dopamine: insights from Parkinson's disease. *Neuron*, 61, 502–510.

de Araujo, I., Rolls, E., Velazco, M., Margot, C., & Cayeux, I. (2005) Cognitive modulation of olfactory processing. *Neuron*, 46, 671–679.

Diener, E., & Chan, M. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being*, 3, 1–43.

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, 22, 780–793.

Dunham, Y., Baron, A.S., & Carey, S. (2011). Consequences of “minimal” group affiliations in children. *Child Development*, 82, 793–811.

Egan, L., Bloom, P., & Santos, L. (2010). Choice-induced preferences in the absence of choice: Evidence from a blind two choice paradigm with young children and capuchin monkeys. *Journal of Experimental Social Psychology*, 46, 204–207.

Egan, L., Santos, L., & Bloom, P. (2007). The origins of cognitive dissonance: Evidence from children and monkeys. *Psychological science*, 18, 978–983.

Ellingsen, D., Leknes, S., & Kringelbach, M. (2015). Hedonic value. In T. Brosch & D. Sander (eds.), *Handbook of Value*. Oxford University Press.

Enisman, M., Shpitzer, H., & Kleiman, T. (2021). Choice changes preferences, not merely

reflects them: A meta-analysis of the artifact-free free-choice paradigm. *Journal of Personality and Social Psychology*, 120, #16.

Fernqvist, F. & Ekelund, L. (2014). Credence and the effect on consumer liking of food – A review. *Food Quality and Preference*, 32, 340–353.

Firestone, C. & Scholl, B. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, 39, e229.

Fodor, J. (1983). *The Modularity of Mind*. MIT Press.

Forsberg, J., Martinussen, M., & Flaten, M. (2017). The placebo analgesic effect in healthy individuals and patients: A meta-analysis. *Psychosomatic Medicine*, 79, 388–394.

Frisaldi, E., Carlino, E., Zibetti, M., Barbiani, D., Dematteis, F., Lanotte, M., Lopiano, L., & Benedetti, F. (2017). The placebo effect on bradykinesia in Parkinson's disease with and without prior drug conditioning. *Movement Disorders*, 32, 1474–1478.

Frisaldi, E., Shaibani, A., & Benedetti, F. (2020). Understanding the mechanisms of placebo and nocebo effects. *Swiss Medical Weekly*, 150, 20340.

Gat, A. (2015). Proving communal warfare among hunter-gatherers: The quasi-Rousseauan error. *Evolutionary Anthropology*, 24, 111–126.

Geers, A., Kosbab, K., Helfer, S., Weiland, P., & Wellman, J. (2007). Further evidence for individual differences in placebo responding: An interactionist perspective. *Journal of Psychosomatic Research*, 62, 563–570.

Gilbert, D. & Wilson, T. (2007). Prospection: Experiencing the future. *Science*, 317, 1351–1354.

Goebel, M. U., Trebst, A. E., Steiner, J., Xie, Y. F., Exton, M. S., Frede, S., Canbay, A., Michel, M. C., Heemann, U., Schedlowski, M. (2002). Behavioral conditioning of immunosuppression is possible in humans. *FASEB Journal*, 16, 1869–1873

Golovanova, L., Doronichev, V., Doronicheva, E., Sapega, V., & Shackley, M. (2021). Long-distance contacts and social networks of the Upper Palaeolithic humans in the North-Western Caucasus (on data from Mezmaiskaya Cave, Russia). *Journal of Archaeological Science: Reports*, 39, #103118.

Grabenhorst, F., Rolls, E., & Bilderbeck, A. (2008). How cognition modulates affective responses to taste and flavor: Top-down influences on the orbitofrontal and pregenual cingulate cortices. *Cerebral Cortex*, 18, 1549–1559.

Gregg, A., Seibt, B., & Banaji, M. (2006). Easier done than undone: Asymmetry in the

malleability of implicit preferences. *Journal of Personality and Social Psychology*, 90, 1–20.

Gross, J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, 26, 1–26.

Guo, J., Yuan, X., Sui, F., Zhang, W., Wang, J., Luo, F., & Luo, J. (2011). Placebo analgesia affects the behavioral despair tests and hormonal secretions in mice. *Psychopharmacology*, 217, 83–90

Hackel, L., Zaki, J., & Van Bavel, J. (2017). Social identity shapes social valuation: Evidence from prosocial behavior and vicarious reward. *Social Cognitive and Affective Neuroscience*, 12, 1219–1228.

Henrich, J. & Gil-White, F. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior*, 22, 165–196.

Henrich, J. (2016). *The Secret of Our Success*. Princeton University Press.

Hill, K., Walker, R., Božičević, M., Eder, J., Headland, T., Hewlett, B., Hurtado, A.M., Marlowe, F., Wiessner, P., & Wood, B. (2011). Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science*, 331, 1286–1289.

Hill, K., Wood, B., Baggio, J., Hurtado, A., & Boyd, R. (2014). Hunter-gatherer inter-band interaction rates: Implications for cumulative culture. *PloS one*, 9, e102806.

House, B., Kanngiesser, P., Barrett, H.C., Broesch, T., Cebioglu, S., Crittendon, A., Erut, A., Lew-Levy, S., Sebastian-Enesco, C. Marcus Smith, A., Yilmaz, S., & Silk, J. (2020). Universal norm psychology leads to societal diversity in prosocial behavior and development. *Nature Human Behavior*, 4, 36–44.

Hrdy, S. (2009). *Mothers and Others*. Harvard University Press.

Inzlicht, M., Shenhav, A., & Olivola, C. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, 22, 337–349.

Izard, C. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, 2, 260–280.

Johnson, A.W. & Gallagher M. (2011). Greater effort boosts the affective taste properties of food. *Proceedings of the Royal Society B,* 278, 1450–1456

Kern, A., Kramm, C., Witt, C., & Barth, J. (2020). The influence of personality traits on the

placebo/nocebo response: A systematic review. *Journal of Psychosomatic Research*, 128, #109866.

Kerner, G., Neehus, A., Philippot, Q., Bohlen, J., Rinchai, D., Kerrouche, N., Puel, A., Zhang, S., Boisson-Dupuis, S., Abel, L., & Casanova, J. (2023). Genetic adaptation to pathogens and increased risk of inflammatory disorders in post-Neolithic Europe. *Cell genomics*, 3, #100248.

Kessner S, Wiech K, Forkmann K, Ploner M, Bingel U. 2013. The effect of treatment history on therapeutic outcome: an experimental approach. *JAMA Intern. Med.* 173:1468–69.

Khan, A., Faucett, J., Lichtenberg, P., Kirsch, I., & Brown, W.A. (2012). A systematic review of the comparative efficacy of treatments and controls for depression. *PLoS ONE*, 7, e41778.

Kinzler, K., Dupoux, E., & Spelke, E. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, 104, 12577–12580.

Kleiman, E., Chiara, A., Liu, R., Jager-Hyman, S., Choi, J., & Alloy, L. (2017). Optimism and well-being: A prospective multi-method and multi-dimensional examination of optimism as a resilience factor following the occurrence of stressful life events. *Cognition and Emotion*, 31, 269–283.

Kobayashi, K., & Hsu, M. (2019). Common neural code for reward and information value. *Proceedings of the National Academy of Sciences*, 116, 13061–13066.

Kudrnova, V., Spelke, E., & Thomas, A. (2023). Infants infer social relationships between individuals who engage in imitative social interactions. *Open Mind: Discoveries in Cognitive Science*, 8, 202–216.

Lee, L.O., James, P., Zevon, E., Kim, E.S., Trudel-Fitzgerald, C., Spiro III, A., Grodstein, F., & Kubzansky, L. (2019). Optimism is associated with exceptional longevity in 2 epidemiologic cohorts of men and women. *Proceedings of the National Academy of Sciences*, 116, 18357–18362.

Liberman, Z., Woodward, A., & Kinzler, K. (2017). Preverbal infants infer third-party social relationships based on language. *Cognitive Science*, 41, 622–634.

Liberman, Z., Woodward, A., Sullivan, K., & Kinzler, K. (2016). Early emerging system for reasoning about the social nature of food. *Proceedings of the National Academy of Sciences*, 113, 9480–9485.

Lieberman, M., Ochsner, K., Gilbert, D., & Schacter, D. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science*, 12, 135–140.

Liu, Y-Z., Wang, Y-X., & Jiang, C-L. (2017). Inflammation: The common pathway in stress-related diseases. *Frontiers in Human Neuroscience*, 11, #316.

Mahajan, N. & Wynn, K. (2012). Origins of "us" versus "them": Prelinguistic infants prefer similar others. *Cognition*, 124, 227–233.

Marlowe, F. (2005). Hunter-gatherers and human evolution. *Evolutionary Anthropology*, 14, 54–67.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., & Sirak, K. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528, 499–503.

McCabe, C., Rolls, E., Bilderbeck, A., & McGlone, F. (2008). Cognitive influences on the affective representation of touch and the sight of touch in the human brain. *Social Cognitive and Affective Neuroscience*, 3, 97–108.

Meeuwis, S., van Middendorp, H., van Laarhoven, A., van Leijenhorst, C., Pacheco-Lopez, G., Lavrijsen, A., Veldhuijzen, D., & Evers, A. (2020). Placebo and nocebo effects for itch and itch-related immune outcomes: A systematic review of animal and human studies. *Neuroscience and Biobehavioral Reviews*, 113, 325–337.

Norton, M., Mochon, D. & Ariely, D. (2012). The IKEA effect: when labor leads to love. *Journal of Consumer Psychology,* 22, 453–460.

Ogilvie, R. & Carruthers, P. (2016). Opening up vision: The case against encapsulation. *Review of Philosophy and Psychology*, 7, 721–742.

Olsson, A., McMahon, K., Papenberg, G., Zaki, J., Bolger, N., & Ochsner, K. (2016). Vicarious fear learning depends on empathic appraisals and trait empathy. *Psychological Science*, 27, 25–33.

Petrie, K. & Rief, W. (2019). Psychobiological mechanisms of placebo and nocebo effects: Pathways to improve treatments and reduce side effects. *Annual Review of Psychology*, 70, 599–625.

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, 107, 8993–8999.

Plassmann, H. & Wager, T. (2014). How expectancies shape consumption experiences. In S. Preston, M. Kringelbach, & B. Knutson (eds.), *The Interdisciplinary Science of Consumption*. MIT Press.

Plassmann, H. & Weber, B. (2015). Individual differences in marketing placebo effects: Evidence from brain imaging and behavioral experiments. *Journal of Marketing Research*, 52, 493–510.

Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences*, 105, 1050–1054.

Powell, L. & Spelke, E. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, 110, E3965–E3971.

Pun, A., Birch, S., & Baron, A. (2021). The power of allies: Infants' expectations of obligations during intergroup conflict. *Cognition*, 211, #104630.

Rakoczy, H. & Schmidt, M. (2013). The early ontogeny of social norms. *Child Development Perspectives*, 7, 17–21.

Schafer, S., Geuter, S., Wager, T. (2017). Mechanisms of placebo analgesia: A dual-process model informed by insights from cross-species comparisons. *Progress in Neurobiology*, 160, 101–122.

Schedlowski, M. & Pacheco-López, G. (2010). The learned immune response: Pavlov and beyond. *Brain, behavior, and immunity*, *24*(2), 176-185.

Schmidt, L., Skvortsova, V., Kullen, C., Weber, B., & Plassmann, H. (2017). How context alters value: The brain's valuation and affective regulation system link price cues to experienced taste pleasantness. *Nature Scientific Reports*, 7, #8098.

Sharot, T., De Martino, B., & Dolan, R. (2009). How choice reveals and shapes expected hedonic outcome. *Journal of Neuroscience*, 29, 3760–3765.

Sharot, T., Fleming, S.M., Yu, X., Koster, R., & Dolan, R. (2012). Is choice-induced preference change long-lasting? *Psychological Science*, 10, 1123–1129.

Sharot, T., Velasquez, C., & Dolan, R. (2010). Do decisions shape preference? Evidence from blind choice. *Psychological Science*, 21, 1231–1235.

Silver, A., Stahl, A., Loiotile, R., Smith-Flores, A., & Feigenson, L. (2020). When not choosing leads to not liking: Choice-induced preference in infancy. *Psychological Science*, 31, 1422–1429.

Sripada, C. & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind volume 2: Culture and cognition,* Oxford University Press.

Steinkopf, L. (2015). The signaling theory of symptoms: An evolutionary explanation of the placebo effect. *Evolutionary Psychology*, 13, 1474704915600559.

Sterelny, K. (2012). *The Evolved Apprentice*. MIT press.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96–102

Thomas, A., Saxe, R., & Spelke, E. (2022). Infants infer potential social partners by observing the interactions of their parent with unknown others. *Proceedings of the National Academy of Sciences*, 119, e2121390119.

Ting, F., He, Z., & Baillargeon, R. (2019). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences*, 116, 6025–6034.

Trimmer, P., Marshall, J., Fromhage, L., McNamara, J., & Houston, A. (2013). Understanding the placebo effect from an evolutionary perspective. *Evolution and Human Behavior*, 34, 8–15.

Troy, A., Shallcross, A., & Mauss, I. (2013). A person-by-situation approach to emotion regulation: Cognitive reappraisal can either help or hurt, depending on the context. *Psychological Science*, 24, 2505–2514.

Wendt, L., Albring, A., Ober, K., Engler, H., Freundlieb, C., Witzke, O., Kribben, A., & Schedlowski, M. (2013). Placebo-induced immunosuppression in humans: role of learning and expectation. *Brain, Behavior, and Immunity*, 29, S17.

Westbrook, A., Lamichhane, B., & Braver, T. (2019). The subjective value of cognitive effort is encoded by a domain-general valuation network. *Journal of Neuroscience*, 39, 3934–3947.

Whitehead, H., Laland, K., Rendell, L., Thorogood, R., & Whiten, A. (2019). The reach of gene-culture coevolution in animals. *Nature Communications*, 10, 2405.

Whiten, A. (2021). The burgeoning reach of animal culture. *Science*, 272, eabe6514.

Wiesmann, C., Kampis, D., Poulsen, E., Schüler, C., Duplessy, H., & Southgate, V. (2022). Cognitive dissonance from 2 years of age: Toddlers', but not infants', blind choices induce preferences. *Cognition*, 223, #105039.

Wiessner, P. (2005). Norm enforcement among the Ju/'hoansi bushmen. *Human Nature*, 16(2), 115–123.

Yang, X., Yang, F., Guo, C., & Dunham, Y. (2022). Which group matters more: The relative strength of minimal vs. gender and race group memberships in children's intergroup thinking. *Acta Psychologica*, 229, #103685.

Zunhammer, M., Ploner, M., Engelbrecht, C., Bock, J., Kessner, S., & Bingel, U. (2017). The effects of treatment failure generalize across different routes of drug administration. *Science Translational Medicine*, 9, eaal2999.