August, 2018

Examining memory for ritualized gesture in complex causal sequences

Kapitány, R[1]., Kavanagh, C[1,2]., Whitehouse, H[1]., & Nielsen, M[3,4].

1. The University of Oxford
2. Hokkaido University
3. The University of Queensland
4. University of Johannesburg

Corresponding Author: Dr. Rohan Kapitány - rohankapitany@gmail.com

School of Anthropology and Museum Ethnography

51/53 Banbury Road

Oxford

OX2 6PE

Accepted for publication in the journal *Cognition*.

Abstract

Humans have created and maintained an exponentially large and sophisticated behavioral corpus over evolutionary time. In no small part this was achieved due to our tendency to imitate behaviours rather than to emulate outcomes. This tendency, however, can lead to inefficiency and redundancy in our behavioral repertoires. Drawing on evidence from multiple fields of psychology, we propose two novel competing hypotheses. The 'catalyst hypothesis' suggests that low (but not high) proportions of ritualized gesture in instrumental action sequences will improve subsequent recall of the entire action sequence (without itself enhancing the instrumental utility of the sequence). Conversely, the 'cost hypothesis' suggests that increasing proportions of ritualized gesture will impair recall, due to the introduction of cognitive load. The null hypothesis states that ritualized gestures are neither beneficial nor costly. In a pre-registered experiment, we presented participants with multiple versions of two complicated 2-minute action sequences in which we varied the proportion of ritualized gesture. We then quantified the influence ritualized gesture had on recall for individuals gestures, overall outcomes, and described detail. We found clear evidence that high proportions of ritualized gestures impair recall for individual gestures and overall success, and weak evidence that low proportions increase overall success. At present, we may reject the null, but cannot rule out either of our competing hypotheses. We discuss potential implications for cultural evolution, and generate competing predictions that allow for adjudication between Ritual Modes theory (Whitehouse, 2004) and the 'Cognitive Resource Depletion' account of Religious Interaction (Schjoedt et al., 2013). All files (including data and syntax) are freely available at https://osf.io/spz68/.

Keywords: Ritual, Memory, Cultural Evolution, Transmission, Culture, Predictive Coding.

1 Introduction

"Make an eyesalve against a wen: take equal amounts of cropleac [an Allium species] and garlic, pound well together, take equal amounts of wine and oxgall, mix with the alliums, put this in a brass vessel, let [the mixture] stand for nine nights in the brass vessel, wring through a cloth and clarify well, put in a horn and at night apply to the eye with a feather; the best medicine." - Bald's Leechbook (Voth, 2017), as described in Harrison et al., 2015)

In 2015 a team of chemists, biologists, and historians attempted to exactly re-create a 1,100 year old medicine for the treatment of a Sty (an infection of the eyelash follicle; Harrison et al., 2015). Given the ritualistic features of the recipe it was to their surprise, that not only was the treatment efficacious, but it was efficacious against MRSA - a particularly problematic strain of treatment-resistant bacteria. The roles played by some ingredients seemed intuitive - garlic, onion, and brass are known to have antimicrobial properties - while the role of other ingredients was less clear. What role did wine play, for example? Did it contribute antimicrobial qualities, or did it merely act as a solvent (p. 2)? Similarly, some of the processes seemed clear and important - pounding and mixing are, in principle, causally relevant - while other processes are opaque. Did the mixture really need to sit nine days, be mixed in a brass vessel, and applied at night with a feather? The team found that a 9-day latency was superior to a 5-day latency, but the use of the brass vessel was irrelevant. These results are surprising, not least because a decade earlier Brennessel, Drout, & Gravel, (2005) concluded, with reference to the same recipe, that *"some of the Anglo-Saxon recipes take biologically efficacious ingredients and process them into ineffective mixtures"* (pp. 184)

One reason it is difficult, a priori, to determine whether such sequences will be effective or not is due to causal opacity: causal opacity can be attributed to an action or a procedure when it is difficult or impossible to determine the physical-causal relationship between the action and the outcome (Kapitány & Nielsen, 2015, 2016; Legare, Wen, Herrmann, & Whitehouse, 2015; Kristoffer Nielbo & Sørensen, 2015; Watson-Jones, Legare, Whitehouse, & Clegg, 2014; Wen, Herrmann, & Legare, 2016). Causal opacity occurs when the causal dependency structure of ordinary actions is disrupted (Schjoedt et al., 2013). Ritualized gestures are causally opaque (and frequently goal demoted; see box 1) due to repetition, redundancy, formality, and/or stereotypy of action (Kapitány & Nielsen, 2015, 2016) - features which disrupt and obscure causal interpretations[1]. Causally opaque actions in context (e.g., the nine day latency, the horn, and the feather) may be causally necessary or entirely arbitrary, the point is that the veridical causal relationship cannot be easily determined.

Ritualistic actions may become embedded, or maintained, within instrumental sequences fairly often (Cohen & Bacdayan, 1994; Keren, Boyer, Mort, & Eilam, 2013). What are the consequences of this? We propose

---

[1] Importantly, ritualistic actions should not be confused with rituals, per se. That is, we make no claim that our stimuli, and the phenomenon under consideration, are ritualistic in the sense that they are symbolic (Durkheim, 2012), or culturally shared (Whitehouse, 2004), merely that they are a set of identifiable actions conforming to the given definition.

here that the inclusion of ritualistic elements in action sequences has 'fitness consequences' for the sequence itself, as the cognitive cost associated with unnecessary and redundant actions may influence the overall quality of recall (particularly as the sequence gets longer and more complicated). We propose two hypotheses, one in which we expect to see a curvilinear relationship between ritualized gesture and recall, such that a small proportion of ritualized gestures will *improve,* rather than *impair* recall (while high proportions will be detrimental), and a second, in which any non-zero proportion of ritualized action leads to impaired recall. As will be described, these effects may be contingent upon a range of cognitive factors, such as predictive coding, executive function, and memory.

Ritualized gestures, particularly in cultural and community rituals, are regarded as costly, requiring expenditure that seems disproportionate to the ostensible outcomes of the act (Henrich, 2009; Ruffle & Sosis, 2007, 2003; Sosis & Bressler, 2003). The same is true at smaller scales, when an ordinary and instrumental sequence contains degrees of redundancy (as in recipes, or procedures for making an artefact, particularly when such procedures are important or hazardous; Boyer & Liénard, 2006; Keren et al., 2013). It is important to note that we use the term 'costly' to refer to the inefficiency of ritualized gesture with regard to the *additional cognitive load* required to remember and reproduce something instrumentally unnecessary (Rybanska, McKay, Jong, & Whitehouse, 2017) - it is not an equivocation between the physical, emotional, financial, and/or opportunity cost associated with participation in large communal rituals. Do we see ritualized gestures within sophisticated action sequences because they are simply not costly enough to be selected against over time and iterations? Or might their presence be explained by a benefit that offsets their costs? Or, finally, are ritualized gestures sufficiently costly that they are selected against and eliminated above a certain threshold? To the extent that ritualized gestures are costly, then their inclusion may represent something of a handicap for cultural learners who have - over evolutionary time - relied on pedagogy, observation, and oral culture to acquire behavioral skills (Cavalli-Sforza & Feldman, 1981; Hewlett, Fouts, Boyette, & Hewlett, 2011; Kline, 2014; Lozada, Ladio, & Weigandt, 2006).

In a rational world, we would evaluate the actions we wish to learn based on the efficacy and efficiency of the technique itself, omitting what doesn't work, selecting that which does, and updating our behavior when we observe a better alternative. But this is not what happens. Rather, we acquire things as a function of learning strategies that may only be proximally related to efficiency or efficacy. In part, this is due to the fact that as our behavioral and technological repertoire became increasingly sophisticated over historic time, it also became increasingly difficult to causally evaluate each individual technique (Legare & Nielsen, 2015). Thus, humans use a range of strategies to acquire important behaviors. Such strategies involve copying *individuals* who are successful or competent (Mesoudi, 2008) particularly when they belong to the in-group (Over & Carpenter, 2012), and using a wider range of models when in a larger group (Kempe & Mesoudi, 2014). And yet, we're not perfect: adults may use prestige information and success-related information equally, even if the former is less useful than the latter (Atkisson, O'Brien, & Mesoudi, 2012), and children may prefer to copy adults rather than age-mates, even if those age-mates have demonstrable competence (Wood et al., 2016; Wood, Kendal, & Flynn, 2012). On the whole, such

strategies are excellent at arriving at *pretty good* solutions (p. 349; Gigerenzer & Selten, 2002), relative to the difficulty and cost of arriving at *optimal* solutions. That is not to argue, however, that certain features aren't cues to efficacy (Barrett & Thomas Lawson, 2001; Legare & Souza, 2012), but rather that the cost of *exclusively* attending to this information is, on average, too high, too difficult, or too costly. The consequence of this is that the strategies we do employ may maintain behaviors which may not be directly, physically-causally, relevant. Such actions, however, may be important. Here we suggest (as others have) that ritualistic actions may cue attention (Rossano, 2012) producing cognitive capture (Boyer & Liénard, 2006) which may facilitate memory and subsequent reproduction.

There is a rich literature on how humans learn motor sequences in order to perform them later (typically referred to as '*delayed motor intentions*'; for review see Badets & Osiurak, 2015). Broadly, there are two dimensions for planning and performing a motor action: 'action-based' and 'goal-based' mechanisms. The former are the [motor] actions required to bring about an effect (e.g., holding and thumbing the flint of a cigarette lighter), while the latter is a conceptualization of the outcome one wishes to bring about (e.g., to generate a flame; Badets, Blandin, & Shea, 2006; Palmer & Meyer, 2000; Schmidt, 1975; Shea & Wulf, 2005). While both of these concepts tend to relate to the individual *performing* the action, they appear to correspond to distinctions made regarding perception, specifically, how actions are causally opaque and/or goal demoted. That is, one may represent the 'action' as a series of motor procedures that corresponds with the learned causal steps of the sequence, while the latter corresponds with an intention to bring about an effect via some means. Both of these are important in understanding how to learn and produce actions, and each alone may be sufficient under some conditions. Importantly, while an 'intention mechanism', and correspondingly, an apparent goal, appear to highly benefit performance, it may primarily be so for skilled actors rather than novices (Palmer & Meyer, 2000). This suggests that domain-novices rely on a causal, or action-based, understanding of each action - thus suggesting that causal opacity of novel ritualized gestures may exploit on a similar mechanism. To the best of our knowledge, this literature has not concerned itself with how actions are learned, represented, or performed when causal and goal information is disrupted or obfuscated (as is the case here). Nonetheless, we will draw upon this literature when interpreting the data and results presented hereafter. However, such questions have been discussed, somewhat orthogonally in the literature on developmental and comparative psychology. There we find a well-developed understanding of how the qualities of such actions influence holistic performance.

Actions which are ritualistic and causally opaque tend to generate in observers a drive to imitate (to recreate the actions) rather than to emulate (to recreate the outcome; Acerbi & Tennie, 2016; Horner & Whiten, 2005; Lyons, Young, & Keil, 2007; Nielsen, Kapitány, & Elkins, 2015; Nielsen, Tomaselli, & Kapitány, 2018). This is commonly referred to as *over-imitation*: the tendency for humans to copy with high-fidelity entire modelled sequences, even when such sequences contain obviously redundant, ineffective, and/or costly actions (Horner & Whiten, 2005; Nielsen & Tomaselli, 2010; Whiten, McGuigan, Marshall-Pescini, & Hopper, 2009). This is despite the fact that children can identify that the very same actions are '*silly*' (Lyons et al., 2007), and may have no causal

importance whatsoever (Keupp, Behne, Zachow, Kasbohm, & Rakoczy, 2015; Nielsen & Blank, 2011). This phenomenon appears unique to humans (Horner & Whiten, 2005; Tennie, Call, & Tomasello, 2006; Whiten et al., 2009), and has been replicated many times over among children and adults (McGuigan, Makinson, & Whiten, 2011; Whiten et al., 2016), in a diversity of varied and unrelated cultures (Berl & Hewlett, 2015; Nielsen, Mushin, Tomaselli, & Whiten, 2014; Nielsen & Tomaselli, 2010) and under a variety of conditions (Keupp et al., 2015; McGuigan et al., 2011; Wood et al., 2016). In their seminal paper, Horner & Whiten (2005) suggest that [over-]imitation "...*requires only that the observer reproduce the actions … with sufficient fidelity to recreate the desired outcome, without having to fully appreciate the causal relationships involved".* We extend this claim in a small but significant way, arguing that 'we are predisposed to reproduce the actions of the model with *increased* fidelity, *because* we cannot fully appreciated the causal relationships involved'. But what does it mean that we cannot *'appreciate the causal relationships'*? How do humans discriminate between ordinary (causally transparent) actions and ritualized (causally opaque) actions? To address this we must first understand how ordinary actions are perceived.

Box 1. *Key terms and definitions.*

> **Ritualized gesture:** *Ritualized gestures are gestures that are causally opaque and frequently goal demoted, that are constituted by repetitive, redundant, stereotyped features, and formality of performance. Ritualized gestures are distinct from rituals, inasmuch as such gestures are often a part of rituals, but not themselves necessarily symbolic and/or culturally shared.*
>
> **Causal Opacity**: *Causal opacity can be attributed to an action or a procedure when it is difficult or impossible to determine the physical-causal relationship between the action and the outcome*
>
> **Goal Demotion.** *Goal demotion can be attributed to an action or a procedure when it is difficult or impossible for the observed to determine the goal or intention of the actor.*
>
> **Overimitation:** *The tendency for humans to copy with high-fidelity entire modelled sequences, even when such sequences contain obviously redundant, ineffective, and/or costly actions*
>
> **Partonomics:** *The study of action perception with regard to how we parse what we see into meaningful units and subunits of information based on multiple levels of prediction and observation*
>
> **Predictive Coding:** *Predictive coding Is a cognitive model of perception in which the brain predicts actions based on sensory input from a top-down perspective, while actively updating these predictions (via subsequent input) when affirmed or violated via bottom-up error-checking processes.*

When we observe others performing actions we parse what we see into meaningful units and subunits of information based on multiple levels of prediction and observation (Gold, Zacks, & Flores, 2017; Newtson, 1973; Nielbo & Sørensen, 2015; Nielbo & Sørensen, 2011; Zacks, 2004; Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001). The study and conceptualization of parsing and partitioning is the study of 'partonomics'. Zacks and colleagues suggest we partition actions, and action sequences, into an hierarchical partonimic structure (i.e., first light the match, then light the cigarette; Zacks, 2004; Zacks & Tversky, 2001; Zacks et al., 2001). There are three basic levels of action parsing (Csibra, 1993; Hamilton, de C. Hamilton, & Grafton, 1993; Zacks, 2004; Zacks &

Tversky, 2001; Zacks et al., 2001). The 'lowest' level, *a gesture*, conforms to one's understanding the discrete motor-actions (i.e., grasping a match). The 'mid' level, *a behavior,* refers to sub-goals and proximal intentions, and is contingent upon our understanding of the causal process and goals that bring about end-states (i.e., striking a match against the box in order to ignite it). While the 'highest' level, *the schema*, refers to the overall intention of the sequence, which itself is contingent upon our familiarity of what we are observing (i.e., smoking a cigarette[2]; Zacks, 2004; Zacks & Tversky, 2001; Zacks et al., 2001)*.* An important part of our understanding of ordinary actions is our ability to make predictions about what we are seeing across levels - this is the study of 'predictive coding' (Huang & Rao, 2011; Kilner, Friston, & Frith, 2007). Broadly put, the brain predicts actions from a top-down perspective, while actively updating these predictions when affirmed or violated via bottom-up error-checking processes. This occurs at each level - gesture, behavior, and schema - but since our predictions are top-down, they are contingent upon our ability to discern superordinate goals and intentions. That is, the demarcation of action units within a sequence is built around the largest discernable goal of the sequence (i.e., the action unit 'lighting a match' is subsumed by the superordinate goal of 'smoking a cigarette'; Wilder, 1978a, 1978b; Zacks, 2004), indicating that there are multiple levels of goal understanding operating simultaneously. Thus, we understand ordinary actions as a series of causal-relationships enacted in order to bring about specific, superordinate, goal-driven, outcomes.

Ritualized gestures (and in particular, *novel* ritualized gestures), because they are causally opaque frequently violate our typical predictive processes. And when a prediction is violated, at any level, it becomes salient and attention-grabbing, producing 'cognitive capture' (Boyer & Liénard, 2006; Liénard & Boyer, 2006; Kristoffer Nielbo, Schjoedt, & Sørensen, 2012; Zacks, 2004; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Ritualized gestures are not easily integrated into a superordinate *schema.* Thus, a sequence that would otherwise be understood as a schema, is regarded as a series of *behaviors,* and behaviors without causal understanding and goal directedness are reduced to *gestures* - simple motoric representations. Once such a violation is noticed we are highly motivated to reconcile 'upwards' and restore more sensible hierarchies (Bekkering, Wohlschläger, & Gattis, 2000; Gleissner, Bekkering, & Meltzoff, 2000; Malle, Moses, & Baldwin, 2001).

But what does it mean that [ritualistic] actions are parsed as *gestures* rather than behaviors or schemas? We may consider larger actions sequences as constituted by 'units' of action. As in the cigarette-lighting example, there is a partonomic demarcation between the holding of the match, and the striking of the match, despite the fact that the motor action may be fluid and continuous. The demarcation between each action, known as an 'event boundary' (Radvansky, 2012), appears to interact with memory in interesting and beneficial ways (Flores, Bailey, Eisenberg, & Zacks, 2014, 2017; Gold et al., 2017). Swallow, Zacks, & Abrams (2009) have found that objects presented at boundaries (between action units) are better remembered than objects presented inside those boundaries. While, Gold et al. (2017) found that introducing artificial boundaries at unnatural junctions increases

---

[2] Indeed, the schema of 'smoking a cigarette' can involve entirely different gestures and behaviors, such as grasping a lighter, and thumbing the flint in order to ignite the fuel.

memory above control condition, though to a lesser extent than boundaries at natural junctions. Indeed, memory effects for attending to event boundaries are apparent even up to one month after viewing stimuli (Flores et al., 2017). Swallow et al (2009) suggest that objects at event boundaries are better remembered due to additional relational and associative processing (p. 12), and this effect is relevant for both short- and long-term memory (p. 21). Convergent evidence shows that deliberately cueing attention to event segments improves overall recall (Gold et al., 2017) and that failure to accurately segment sequences (among cognitively impaired older adults) predicts omission errors when reproducing action sequences (Bailey, Kurby, Giovannetti, & Zacks, 2013). Indeed, the consistency with which an individual agrees with the consensus of an event boundary's location predicts memory for the event itself, and does so uniquely and independently from the overall knowledge of the superordinate event (Gold et al., 2017). Thus, if the inclusion of ritualistic actions within larger action sequences generates more event boundaries, and event boundaries increase specific and overall recall, then it may be the case that ritualistic actions, while cognitively costly, may convey a memory advantage for the sequence itself (Sargent et al., 2013).

Of course, our capacity to make predictions and detect errors, or to simply learn something, is not free. The cognitive load associated with detecting errors appears to lead to deteriorating performance on key measures (Botvinick, Cohen, & Carter, 2004; Carter, Botvinick, & Cohen, 1999; Engle, 2002), particularly when more event boundaries are present (Kristoffer Nielbo et al., 2012; Kristoffer Nielbo & Sørensen, 2015). Schjoedt et al. (2013) have argued that rituals deplete executive resources specifically as a function of their goal demotion and causal opacity (see also: Rybanska et al., 2017). Under extreme circumstances, as with high arousal 'imagistic' rituals (Whitehouse, 2004), chronic depletion produces an 'inferential gap' that can be resolved at a later date with orthodox explanations (Schjoedt et al., 2013) or with self-generated meaning-making processes (Richert, Whitehouse, & Stewert, 2005; van Mulukom, 2017; Whitehouse, 2004). However, the present analyses relate to ritualized gestures, rather than rituals per se (though such gestures are almost invariable present in rituals proper, and are relevant to ritual theories). Moreover, the following experiments take place under circumstances when the observer is naive to the sequence, akin to a pedagogical situation in which one must learn the sequence for the first time. We predict that when ritualised actions are too frequent, the learner's executive resources are depleted to such an extent that memory for the overall action sequence is diminished. Thus, it may be the case that there is an optimal frequency of embedded ritualized gestures within larger action sequences, and this frequency has consequences for the transmission of the sequence between models and learners.

To summarize, ritualized gestures violate our predictions and causal understanding of what we observe. This has multiple downstream consequences. First, ritualized gestures create more event boundaries (which appear to benefit our memory). Once detected, they motivate us to restore an hierarchical, schematic, understanding the of the actions we observe, and yet, the cognitive costs associated with additional attention and reconciliation of prediction errors increases, potentially leading to impairment above a certain threshold. Thereafter, perhaps as a consequence of our ability to recognize ritualistic actions, we are motivated to imitate,

rather than to emulate them. Thus, we may return to our key question: *does the inclusion of ritualistic elements in large and/or instrumental sequences have 'fitness consequences' on the recall and transmission of the sequence itself [under initial learning conditions]?* This question has implications for understanding why ritualized gestures seem to persevere in some contexts. If they benefit memory (per one prediction) this may aid naive individuals in acquiring the whole sequence (at least during the initial stages of acquisition). For any given action sequence, the presence of ritualized gesture may be a '*good trick*' (Dennett, 1996; Weber & Depew, 2003) that enhances recall for the entire sequence. On the other hand, ritualized actions may simply be costly, affording no benefit to a learner, and persevering for other reasons. Thus, we describe three hypotheses:

Hypothesis 1 - The 'Catalyst Hypothesis: Ritualistic actions are cognitively costly (they deplete cognitive resources), but this cost is offset by a cognitive benefit (i.e., memorability, 'deeper' processing) under specific conditions. When the proportion of ritualistic actions within a sequence is sufficiently low, the benefit is greater than the cost, and performance on recall measures is improved above baseline performance. When the proportion is high, performance declines below baseline measures. Performance will thus be approximately described by an inverted-U shaped function (peaking at low proportions), and declining as the proportions of ritualistic actions increase.

Hypothesis 2 - The 'Cost' Hypothesis: Ritualistic actions are costly, and no benefit exists to offset them. As the proportion of ritualistic actions increase, performance on recall measures will continue to decline. Performance will be described by an overall negative function declining from baseline (i.e., the intercept).

Null Hypothesis: Ritualistic actions are neither cognitively costly, nor beneficial. There will be no clear impact on recall (across various measures) as the proportion of ritualistic actions increase within a given action-sequence. Performance will be described as a flat, linear function as ritualistic actions increase.

That which follows is a report and discussion of the replication of a pre-registered experiment. The initial pre-registered study, reported in full in Supplementary Material A, was potentially compromised due to an oversight in the design. The following study is a replication, employing a more refined design, while retaining core predictions and stimuli.

2 Experimental Design

2.1 Stimuli and design

We created two complex 120-second recipe-like action sequences, each of which consisted of six sub-goals that needed to be satisfied in order to produce a final product. Two 'master sequences' were created, which exclusively featured instrumental action (there were no ritualistic features; 'Sequence-0'). Six variations were

made of each master-sequence, where increasing proportions of the sub-goals were made ritualistic (see Figure 1 for operationalization). The first variation (Sequence-1) featured one ritualistic action and five instrumental actions, the second variation (Sequence-2) features two ritualistic actions, and four instrumental actions, and so on. The sixth and final sequence (Sequence-6) featured only ritualistic actions. In total, there were seven variations for each master sequence. We provided participants with the cover-story that both sequences were medieval recipes, where one sequence produced a drink for the treatment of *worms* (a gut parasite), and the other sequence produced a paste for the treatment of a *burn*. Table 1 describes each sequence, and what the instrumental and ritualistic actions were in each. It is worth noting that when a ritualistic act is first presented (in, say, sequence-1), that ritualistic action is preserved in all subsequent variations (in sequence-2, -3, -4, etc.). Put another way, this experiment did not control for *where* in the sequence a ritualistic action was performed, but manipulated the *frequency* of ritualistic action in the sequences overall. Links to all videos are in Supplementary Material B.

Unlike the initial study (reported in Supplementary Material A) only four of the seven variations were used, corresponding to distinct conditions. These conditions are: Sequence-0, Sequence-2, Sequence-4, and Sequence-6. All participants completed one of the two Sequence-0 videos as a baseline measure. Thereafter, they were randomly assigned to one of the four possible videos from the alternate sequence.

Figure 1. *Operationalizing ritualized gestures from instrumental gestures*
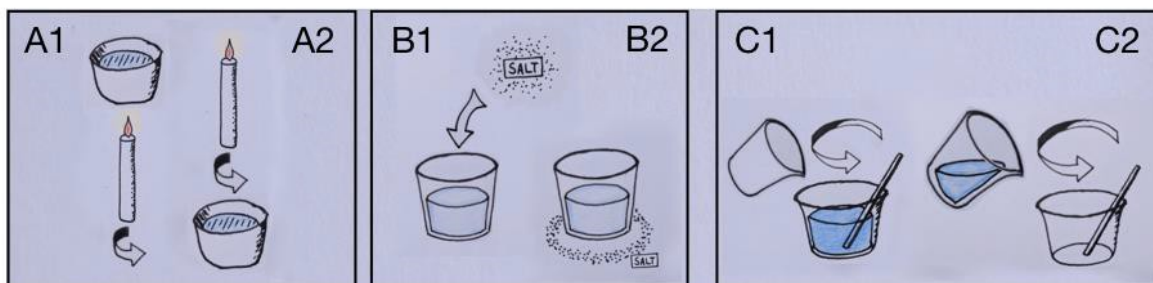
Note: *In A1 the candle heats the water, in A2 the waved above the water. In B1 the salt goes into the water, in B2 the salt is sprinkled around the water. In C1 the rod stirs the water, in C2 the rod stirs the empty bowl. In each instance an intuitive causally transparent and goal apparent action is rendered causally opaque and goal demoted.*

2.2 Data Collection

We aimed to recruit 400 individuals via Amazon's mTurk platform. This corresponds to 100 participants per cell. Simulated data, based on means and standard deviations obtained in the initial study (reported in Supplementary Material A), indicated that 100 participants per conditions would be sufficient to detect a 3% change in dependent variables between contiguous conditions. Participation lasted between 20 and 30 minutes.

Participants were awarded $5US. In total, we collected 412[3] responses. The mean age was 35.06 years (*SD* = 9.43), of whom 177 (43.00%) were female, 233 (56.60%) were male, and 2 (0.5%) did not specify their sex.

---

[3] It is sometimes possible to over-collect on mTurk if participants do not officially conclude their session.

Table 1. *The constituent actions of both master sequences, with instrumental and ritualized variations presented side-by-side.*

Worms Sequence		
'Behavior'	**Instrumental Gesture**	**Ritualized Gesture**
1	Open a jar Spoon garlic into jar Put a pinch of rosemary into jar Screw on lid Shake Jar (5 seconds) Put Jar Down	Open a jar Spoon garlic into jar Put a pinch of rosemary into jar Screw on lid Put Jar Down Shake Hands (5 seconds)
2	Put a shotglass/measure of oil into new jar Put a shotglass/measure of milk into the same jar Stir with a rod (3 seconds) Tap dry rod	Put a shotglass/measure of oil into new jar Put a shotglass/measure of milk into the same jar Stir outside the jar with a rod (3 seconds) Tap rod
3	Open the first jar, empty into a cloth Massage the cloth (5 seconds) Place over jar Strain with water Discard waste	Open the first jar, empty into a cloth Massage the cloth (5 seconds) Discard materials pour in water -
4	Light a candle, place on the table Light a candle, place **next** to first Light a candle, place with other two Heat the oil/milk mix over the flame (10 seconds)	Light a candle, place on the table Light a candle, place **away** from the first Light a candle, place **away** from the other two Heat the oil/milk mix over the flame (10 seconds)
5	Pour the oil/milk mix into a shotglass/measure Pour into the water mix Pour the oil/milk mix into a shotglass/measure Pour into the water mix	Pour the oil/milk mix into a shotglass/measure Pour back Pour the oil/milk mix into a shotglass/measure Pour into the water mix
6	put salt into the glass with hands put salt into the glass with hands	put salt around the glass in a half-circle with hands put salt around the glass in a half-circle with hands
Burns Sequence		
'Behavior'	**Instrumental Gesture**	**Ritualized Gesture**
1	Pour a shotglass/measure of milk into a bowl	Stir the empty bowl with a stirring rod (5 seconds)

	Put a pinch of salt into the bowl Stir the mix with a stirring rod (5 seconds)	Pour a shotglass/measure of milk into a bowl Put a pinch of salt into the bowl
2	Take a spoon **by the handle** Measure out a spoon of oil Pour into the mix	Take a spoon **by the bowl** Pour oil onto the spoon … ... into the mix
3	light a candle hold the bowl above the candle heat the bowl (10 seconds)	light a candle hold the bowl **under** the candle maintain (10 seconds)
4	Pour some honey into a shotglass take a pinch of tobacco and roll in hands put in honey take a pinch of tobacco and roll in hands put in honey take a pinch of tobacco and roll in hands put in honey	Pour some honey into a shotglass Rub hands take a pinch of tobacco and put in honey Rub hands take a pinch of tobacco and put in honey Rub hands take a pinch of tobacco and put in honey
5	Heat the honey mix (10 seconds) Scrap out mix with rod into bowl Blow out candle	Blow out candle Hold the honey mix over the candle (10 seconds) Scrape out mix with rod into bowl
6	Add a spoon of flour into the mix mix 5 seconds scrap clean mix 5 seconds Scrap clean	Add a spoon of flour into the mix Stir outside of bowl for 5 seconds scrap mix 5 seconds Scrap clean

Note: *In the ritualized column the grey text denotes actions that were preserved from the instrumental sequence, while the black text in denotes the variations which 'ritualized' the action.*

2.3 Dependent Variables

Dependent Variables correspond to standard measures in the memory literatures (see: (Gold et al., 2017; Koriat, Ben-Zur, & Nussbaum, 1990; Zacks, Speer, Vettel, & Jacoby, 2006).

2.4 Accuracy. 'Accuracy' describes a measure in which participants were correctly able to categorize whether a particular action was performed in the observed sequence. It was measured using 10 true/false statements. In all conditions the 10 questions were the same, however, in some conditions the correct answer was 'true' while in other conditions the correct answer was 'false' (See Table 2 for questions). All 10 questions were presented simultaneously in random order. This measure is insensitive to the order in which the actions actually took place.

2.5 Success. The 'Success' measure describes the degree to which participants were able to correctly order 12 true statements about the actions performed in the observed sequence (See Table 3 for questions in correct corresponding order). All 12 items were presented simultaneously in random order. Participants were

awarded one of two scores: 1 (all items were ordered correctly) or 0 (items were ordered incorrectly). This measure was designed to reflect whether the participant could produce the intended *outcome* of the sequence (hence the binary nature). This measure is entirely contingent upon the correct nesting of operations, and is somewhat robust against issues of recall, as the items were true, and were described to the participant as such.

2.6 Detail. 'Detail' is a measure designed to describe how much detail participants could generally ascribe to the action sequence when asked to recall the sequence freely. It was a simple word count (discussion and validation of why this was chosen is available Supplementary Material A).

2.7 Cognitive Load. We asked participants to respond to the question *"How good was your performance on the [test] sequence?"* (on 0 to 100 point scale, where zero was 'terrible' and 100 was 'excellent, and 50 was 'average'), for each video they observed, after all primary measures had been collected (Koriat et al., 1990; Rosenbaum & Gregory, 2002). This measure, then, is a proxy for cognitive load - that is, people who felt more burdened by a complex sequence would report lower confidence in their performance. We are not explicitly interested in the relationship between veridicality and confidence, though it is the case that when people are incorrect they tend to overestimate how veridical they are (Bornstein & Zickafoose, 1999). What we are interested in is the relationship between these scores across conditions, and not between these scores and actual performance. Further, we asked participants to rate how good their memory was *generally*, both before and after the test phase of this experiment. This was done on the same 100-point scale. This was a control measure - if confidence *generally* remains unchanged over the course of the experiment, and is different from confidence *for a sequence* we can infer that the sequence was responsible for the change.

2.8 Miscellaneous measures. In addition to these variables, participants filled out a brief demographic survey both before and after the experiment. Finally, participants were asked four questions on a 7-point scale (from strongly disagree to strongly agree) on the following statements: *"The final product was physically ordinary, like the product of any other recipe", "The final product was special in a way other physical products are not", "I expect the final product to be an effective medicine for treating [worms/burns]", "The final product had a supernatural quality about it".* We asked these questions in order to follow-up on results obtained by Kapitány & Nielsen (2015, 2016). As they are not central to the primary hypotheses, results of these analyses are reported and discussed in Supplementary Material C.

Table 2. *The True/False questions 'Accuracy' Questions for each of the sequences.*

	Worms	**Burns**
1	The model shook Jar 1	The model stirred the milk and salt together
2	The model stirred the contents of Jar 1	The model measured oil into the cupped end of a spoon
3	The model used the cloth to strain a liquid	The model held the salt/milk mixture above the candle
4	The model placed all the candles very near to each other	The model ground the tobacco into his palm
5	The model added two measures of Oil/Milk into the water	The model heated the tobacco mixture above a flame
6	The model made a ring of salt around the final mixture	The model drew circles around the final mixture with a spoon
7	The model used his fingers to extinguish the candle(s)	The model used his fingers to extinguish the candle(s)
8	The model tasted the final mixture	The model tasted the final mixture
9	The model clapped his hands	The model clapped his hands
10	The model used all the ingredients/items on display at least once	The model used all the ingredients/items on display at least once

Note: The correct response (true or false) varied depending on which condition the participant was in.

Table 3. The statements participants ordered to determine the 'Success' measure.

	Worms	Burns
1	Garlic was added	A shot glass of milk was poured into bowl
2	Rosemary was added	The salt was added
3	A jar was shaken	The stirrer was used to stir
4	Oil and milk were put into a jar	The small spoon was used to measure the oil
5	The stirrer was used to stir the contents of a jar	The candle was lit
6	Water was poured through the cloth	The bowl was held above the candle
7	The contents of the cloth were discarded	Honey was put into the SMALL glass
8	The candle(s) were lit	The model crushed the tobacco
9	The candle(s) were used for heating	The honey/tobacco mixture was held above the candle
10	The oil milk mixture was poured into a shotglass	The candle was extinguished
11	The oil milk mixture was poured into a shotglass (2)	Flour was added to the mixture
12	The salt was used	The big spoon was used for stirring

3 Pre-registration, Hypotheses, and Analyses Plan

We made several pre-registered predictions (pre-registration can be found online here:

https://osf.io/dhy76/registrations/)

1. A small ratio (i.e., sequence-2) of ritualized gestures will significantly improve Accuracy and Success above the fully instrumental sequence (sequence-0).
2. A high ratio (i.e., sequence-6) of ritualized gestures will result in lower Accuracy and Success compared to the fully instrumental sequence.
3. There will be a linear and positive relationship between ratio of ritualized gestures and detail.

We will use a between-subjects ANOVA to evaluate the mean differences in measures of Accuracy (range 0 - 10) of each condition. Predictors include condition, sequence (whether the participants saw a worms or burns

video during the test condition), and their baseline scores on the respective measure. We will use a logistic regression to analyses (binary) Success (1 = No errors present / all 12 items in correct order; 0 = Error presents / Incorrect overall order). Detail / Word count was analysed using a GLM, with condition, sequence, and baseline word count as predictors.

3.1 Results (pre-registered)

  3.2 Accuracy. We conducted a Mixed ANOVA in which accuracy was included as a within-participants variable (Baseline Accuracy and Test Accuracy), and condition and which sequence they observed in the baseline phase as a between participant factors. Age and sex were entered as controls. Values of accuracy for all for conditions analysed were inside acceptable thresholds for skewness (-1.25 - -0.44) and kurtosis (-.32 - 1.72) (George & Mallery, 2011). We did not observe a significant within-participants difference on scores of accuracy, Baseline Accuracy ($M = 8.50$, $SD = 1.46$) was not significantly different from Test Accuracy ($M = 7.83$, $SD = 1.76$), $F(1,402) = 1.82$, $p = .178$, $d = .40$. However, we did observe a significant interaction between accuracy scores and condition, $F(3, 402) = 26.21$, $p < .001$. Simple effects between baseline and test scores for each sequence are presented in Figure 2. These results indicate that participants who observed sequence-4 ($M = 7.35$, $SD = 1.81$), $t(103) = 6.63$, $p < .001$, $d = .74$, *95%CI* 0.815 - 1.512, and sequence-6 ($M = 7.02$, $SD = 1.98$) $t(101) = 9.68$, $p < .001$, $d = 1.17$, *95%CI* 1.240 - 1.878, were less accurate than on their baseline (while no difference was observed for those whose test condition was sequence-0 ($M = 8.60$, $SD = 1.23$), $t(103) = 1.06$, $p = .29$, $d = .10$, *95%CI* -0.443 - 0.135), or sequence-2 ($M = 8.34$, $SD = 1.47$, $t(100) = .92$, $p = .359$, $d = .09$, *95%CI* -0.161 - 0.437).

Figure 2. *Simple effects for the Accuracy by Condition interaction.*
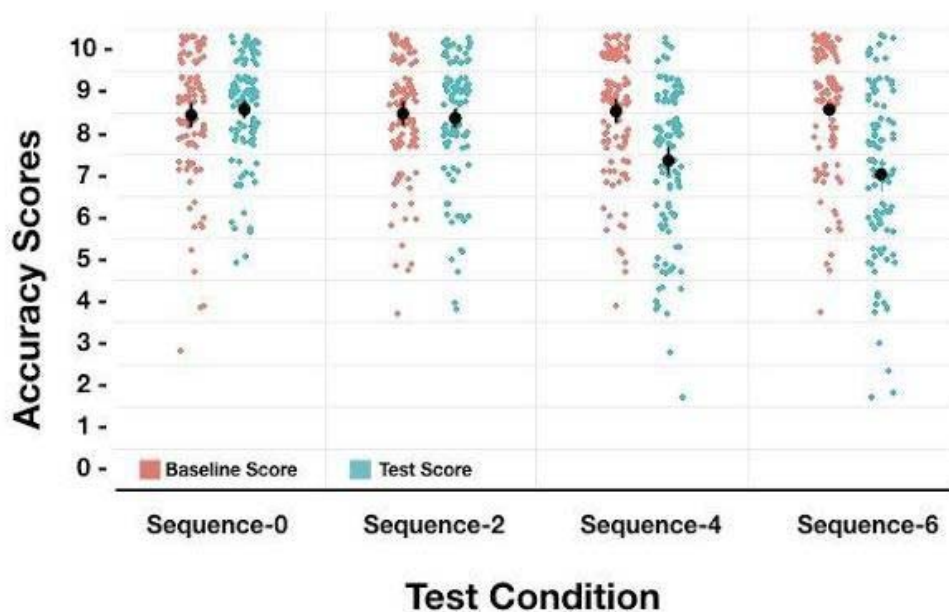
Note: Means and 95%CI shown in black.

On our between-participants analyses, we observed a main effect of condition, $F(3, 402) = 7.80$, $p < .001$. We conducted follow-up analyses with bonferroni correction and found that Sequence-0 ($M = 8.60$, $SD = 1.23$) was not different from Sequence-2 ($M = 8.34$, $SD = 1.47$), $p = .59$, $d = .19$, *95%CI* -0.264 - 0.462, but was significantly different from Sequence-4 (M = 7.35, SD = 1.81), $p = .001$, $d = .81$, *95%CI 0*.232 - 0.951, and Sequence-6 (*M* = 7.02, *SD*= 1.80), $d = .60$, $p < .001$, $d = 1.03$, *95%CI* 0.380 - 1.103. Sequence-2 was significantly greater than Sequence-4, *p* = .008, $d = .60$, *95%CI 0*.131 - 0.855, though sequence-4 was not significantly different from Sequence-6, $p = .415$, d = .17, *95%CI* -0.510 - 0.211. Figure 3 shows a plot of accuracy score per condition with 95% confidence intervals.

Figure 3. *A plot of accuracy scores between conditions.*
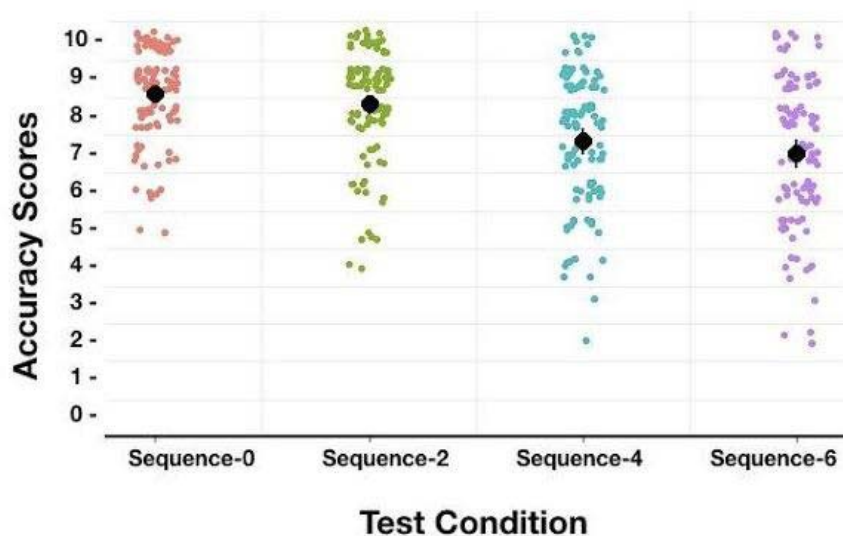
We also found a main effect of sequence observed, $F(1, 402) = 7.41$, $p = .007$, such that those who observed the Worms sequence at baseline produced lower accuracy scores in the test [Burns] sequence ($M = 7.57$, $SD=1.89$) than those who observed the Burns sequence at baseline and the Worms sequence at test ($M = 8.07$, $SD= 1.61$), $p = .007$, $d = 0.28$, *95%CI* -0.610 - -0.098. There was also a significant main effect of age, $F(1, 402) = 4.53$, $p = .034$, $\eta^2_p = .011$, such that as participants get older, their performance improves slightly. There was no main effect of sex, $F(1, 402) = .65$, $p = .419$.

3.3 Success. We can see the relative proportions in Figure 4 for the binary measure of 'Success'. We used a logistic regression, where the dependent variable was Success (1) or Failure (0), and the predictors were baseline success (1/0), the sequence observed, age, and sex. The model had a significant fit, $\chi^2(7) =21.08$, $p = .004$, and explained between .05 (Cox and Snell) and .09 (Nagelkerke) variance in the model. We found that condition was a significant predictor, $\chi^2(3) = 8.15$, $p = .043$, as was Baseline success, $\chi^2(1) = 10.66$, $p = .001$. We found that those who failed the baseline measure were .347 times less likely to succeed in the test measure ($p < .001$, *95%CI* 0.187 - 0.643). We also found that, compared to those who saw Sequence-6 in the test phase, those who saw Sequence-2 in the test phase were 3.218 times more likely to succeed than fail ($p = .007$, *95%CI* 1.376 - 7.522). Neither age ($p = .73$, 95%CI 0.976 - 1.035) nor sex ($p = .727$, *95%CI* 0.629 - 1.943), nor which sequence the participant observed ($p = .278$, 9*5%CI* .781 - 2.365) predicted success.

Figure 4. *The proportion of 'successful' responses per condition.*
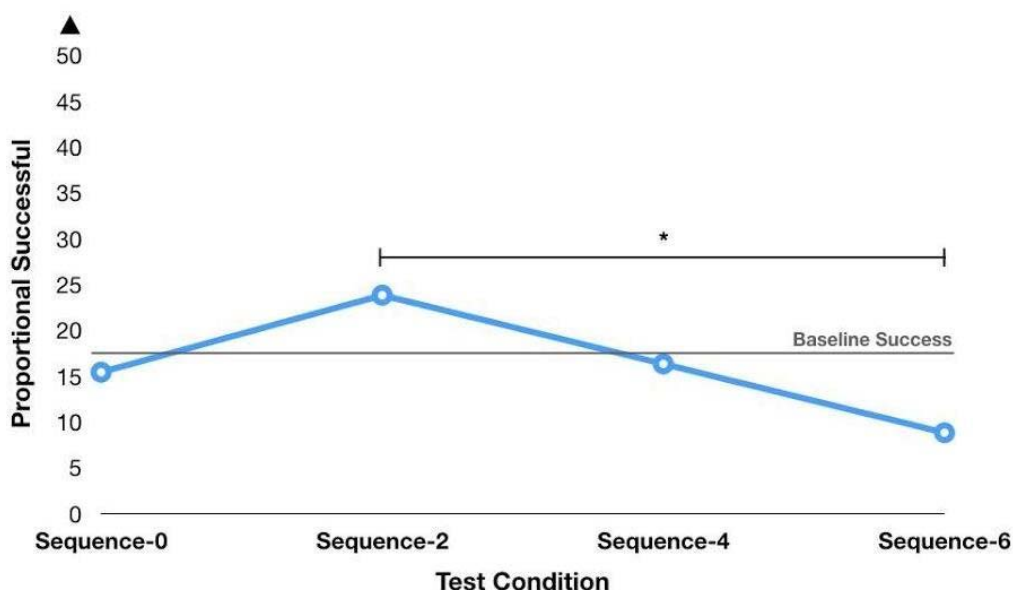
3.4 Detail. We conducted a Mixed ANOVA on detail (word count) in which word count from the baseline sequence was included as a within-participants predictor, while the condition and the sequence observed as between-participant factors, as well as age and sex. We note that the assumption of sphericity was violated (bartlett's test, p = .029), though all corrected values were identical to the uncorrected values, and so 'uncorrected' values are reported here. We did not observe a significant difference in word count between baseline (M = 91.75, SD = 43.47) and the test sequence (M = 93.65, SD = 48.11), $F(1,402)$ = 3.75, p = .054. However, we did observe a significant interaction between word count scores and condition, $F(3, 402)$ = 9.85, p < .001, a significant interaction between word count and which sequence they observed, $F(1, 402)$ = 50.32, p < .001, and a significant interaction between word count and age, $F(1, 402)$ = 5.30, p = .022. Simple effects for these analysis are presented in Tables 4 and 5.

Table 4. *Simple effects (paired sample t-tests) for the Word count * condition interaction.*

Condition	Baseline Word Count	Test Word Count	*d*	Analysis
Sequence-0	93.38 (49.12)	82.58 (42.62)	.37	*t*(103) = 3.90, *p* < .001
Sequence-2	85.79 (38.47)	89.30 (42.87)	.12	*t*(100) = 1.11, *p* = .269
Sequence-4	96.25 (41.06)	99.67 (48.61)	.12	*t*(103) = 1.00, *p* = .287
Sequence-6	91.63 (44.00)	102.87 (54.92)	.41	*t*(101) = 3.51, *p* = .001

Table 5. *Simple effects (paired sample t-tests) for the Word count * sequence interaction.*

Sequence	Baseline Word Count	Test Word Count	*d*	Analysis
Worms	96.45 (44.63)	87.75 (43.59)	.29	*t*(200) = 4.11, *p* < .001
Burns	87.36 (41.78)	99.19 (51.42)	.44	*t*(209) = 5.52, *p* < .001

On measures of detail during the test phase, neither skewness nor kurtosis exceeded acceptable thresholds (skewness: .33 - .78; kurtosis: -.38 - 1.27; George & Mallery, 2011). On between participant measures, we did not observe a main effect of condition, $F(3, 402) = 1.54$, $p = .204$; Sequence-0 ($M = 82.58$, $SD = 42.62$) was not different from Sequence-2 ($M = 89.30$, $SD = 42.87$), $p = .948$, nor sequence-4 ($M = 99.67$, $SD = 48.61$), $p = .108$, nor Sequence-6 ($M = 102.87$, $SD = 54.92$). We did not observe a main effect associated with which sequence they observed, $F(1, 402) = .23$, $p = .634$, nor with participant sex, $F(1, 402) = 3.58$, $p = .059$. There was a main effect of age, $F(1, 402) = 10.51$, $p = .001$, such that older participants wrote more words.

3.5 Non-Pre-registered Analysis. We asked people to rate how good their memory was generally, both before and after the test phase of the experiment. This was done on a scale of 0 ('Terrible') to 100 ('Excellent'), with a mid-point of 50 ('Average'). As expected, there was no difference between conditions on the pre-measure of general ability (collapsed across all condition: $M = 62.16$, $SD = 18.12$), $F(3, 409) = 1.09$, $p = .354$. We conducted a Mixed ANOVA in which confidence in general memory was included as a within-participants variable (pre- and post- measures), with condition and sex as between factors, and age as a covariate. We observed a significant difference between pre- ($M = 62.16$ $SD = 18.12$) and post- measures ($M = 59.80$, $SD = 19.17$), $F(1, 401) = 5.41$, $p = .021$, $d = .22$. However, we observed no effect of condition, $F(3, 401) = 1.25$, $p = .290$, nor an interaction between pre- measure and condition, $F(3, 401) = .90$, $p = .443$. No other main effects or interactions were observed. These results suggests that a participant's experience of the study appeared to diminish their evaluation of their own perception of memory abilities generally, but this was not associated with the condition in which they participated.

Participants were also asked to respond to the question *"How good was your performance on the [test] sequence?"* (on the same 0 - 100 scale). We conducted a linear regression on self-reported performance, with self-reported general memory scores (Pre), condition, sequence observed, and age and gender entered as predictors. The model produced was significant, $F(5, 407) = 13.46$, $p < .001$. Self-reported general memory scores significantly predicted self-reported performance for test-sequences, $β = .359$, $t(402) = 7.75$, $p < .001$, *95%CI 0*.354 - 0.594. Condition also significantly predicted self-reported performance, $β = -.115$, $t(402) = 2.49$, $p = .013$, *95%CI* -2.198 -- 0.258, such that as the number of ritualized gestures increased, self-reported performed declined. The sequence did not significantly predict self-perception of memory performance, $β = -.015$, $t(402) = .75$, $p = .748$, *95%CI* -5.047 - 3.630. Neither age, $β = .053$, $t(402) = 1.13$, $p = .260$, *95%CI* -0.100 - 0.370, nor gender, $β = -.041$, $t(402) = .87$, p .385, *95%CI* -2.498 - 6.467 predicted perceptions of self-reported memory ability for the test-sequences. Descriptively, the patterns of self-reported performance for the test sequences is as follows: Sequence-0: $M = 54.69$, $SD = 24.37$; Sequence-2: $M = 52.51$, $SD = 23.11$; Sequence-4: $M = 49.28$, $SD = 27.19$; Sequence-6: $M = 48.79$, $SD = 23.92$.

4 Conclusions

With regard to our accuracy measure, a set of 10 true/false statements, we found that participants were approximately equally accurate on sequence-0 and sequence-2. Similarly, participants were approximately equally accurate on sequence-4 and sequence-6. However, we found that participants were more accurate on sequence-0 and -2 than on sequence-4 and -6. That is, high proportions of ritual decrease performance relative to baseline, and low proportions did not produce a detectable difference from one's own baseline.

With regard to our success measure, a task that required the ordering of 12 true statements we found that relative to those who observed sequence-6, those who observed sequence-2 were 3.40 times more likely to be successful. As observable in Figure 6, the maxima in overall success was in the predicted 'low proportion' location.

Finally, on our measure of 'detail', a simple word count obtained from a description of the stimuli, we observed no significant effect associated with condition. This data is most consistent with the null hypothesis.

We observed that participants reported declining confidence in their memory abilities in the test condition as the proportion of ritualistic elements increased (while the small change in confidence for general memory appeared uninfluenced by which condition participants were in). This provides suggestive evidence that increasing ritualistic features contributes to an increased cognitive load.

5 General Discussion

In this paper we attempted to quantify the role of ritualized gesture in otherwise instrumental action sequences. Our hypotheses were generated from a synthesis of evidence and theory from a range of disciplines, including cognitive psychology, developmental and comparative psychology, and cultural evolution. Otherwise ordinary and

instrumental gestures were rendered ritualistic – such actions were causally opaque, as the causal relationships between the action and the outcome was not able to be determined (as they simply did not exist).

The 'catalyst hypothesis' – so called because the ritualized gesture does not contribute to the instrumentality of the sequence - predicts that a small degree of ritualized gesture improves the recall for a whole sequence (while a large amount has a negative impact). The 'cost hypothesis' predicts performance will increasingly decline with increasing levels of ritualized gesture. While the null hypothesis predicts no relation. We measured the influence such actions had on three key measures: accuracy (whether participants could correctly identify and/or recall specific actions), success (whether participants could order specific actions in such a way as to exactly reproduce the outcome of the sequence) and detail (a measure of how many words participants used to describe the sequence they observed).

The mechanism of action here is two-fold. First, we assumed, on the weight of evidence (Gold et al., 2017; Newtson, 1973; Kristoffer Nielbo & Sørensen, 2015; K. Nielbo & Sørensen, 2011; Zacks, 2004; Zacks & Tversky, 2001; Zacks et al., 2001), that ritualistic actions introduce an increased number of partonomic 'units' (Kristoffer Nielbo et al., 2012; Schjoedt et al., 2013) , as such actions increase the frequency of perceived event boundaries due to the causally opaque nature of the ritualistic actions. Second, an increased number of ritualistic gestures increase prediction error, thus burdening the observers' cognitive resources. Before summarizing the results, let us consider some criticisms of the mechanism and our stimuli. One criticism is that we neither validated nor quantified the increase in event boundaries / action units as a consequence of the introduction of ritualised gestures. Assuming the claim is correct, that causally opaque gestures do force increased segmentation, we must also assume - at least in the present design - that the exact increase between each step may not be equal. A resultant corollary is that worms *sequence-x* may not have an equal number of boundaries/units as burns *sequence-x*. In the former, we suggest that the exact number is not important at the present stage of research. We are concerned with the general *proportion* of 'ritualistic actions' within an action sequence, and not necessarily the proportion of error-inducing action units (though one necessarily follows from the other). Moreover, since the gestures that were ritualized in conditions were maintained in subsequent conditions, it is the case that (even if each difference between conditions was not equal) the increase in consequent boundaries / units is approximately linear. That we were interested in proportion of ritualised *gestures* speaks to the early stages of testing these hypotheses, while allowing for fine-grained resolution on the task in the future. To that end, all our stimuli and data are available freely online at https://osf.io/spz68/. In the instance of the corollary, it may also be the case that corresponding variations of each sequence may not be equal. The described logic applies (that we are interested in the proportion of ritualized behaviors); we also point out that the design was such that the differences ought to balance over conditions, since participants had the order of variations randomly allocated (and this was statistically accounted for). Though we do concede, in this instance, more fine-grained quantification would have allowed for more power in detecting an effect, our emphasis on the behavioral and cultural consequences of ritualized gesture would mean that such small statistical effects would be uninteresting within the given framework.

With regard to our results, we found that, on a within-participants measure, that those who responded to Sequence-4 and Sequence-6 were less accurate than their baseline measures. There was also a main effect of condition between participants, such that scores on Sequence-0 and Sequence-2 were not statistically different, but scores in Sequence-0 and -2 were significantly more accurate than those reported in Sequence-4 and -6. Put simply, low proportions of ritualized gesture did not appear to influence within- or between- measures of accuracy, but moderate- and high- levels influenced accuracy negatively. The latter finding is consistent with both alternative hypotheses. On measures of accuracy, we may reject the null hypothesis, but cannot discriminate between the two alternatives.

On a measure of 'success' we found a similar pattern. Those who observed sequence-2 were 3.40 times more likely to be correct than those in sequence-6, though we did not observe any other differences. And while we observed the highest rate of success in sequence-2, and the lowest rate in sequence-6, they were not significantly different from scores in sequence-0.

On measures of 'detail', which were the number of words participants used to describe the sequence in as much detail as possible, we observed, on a within-participants measure, that those who observed sequence-0 wrote significantly fewer words than in their own baseline, and significantly more words in sequence-6 (with intermediate values for sequence-2 and -4). This trend is consistent with the interpretation the rituals generate high rates of event segmentation (Kapitany & Nielsen, 2017). There may be some valid criticism of the 'word count' as a measure of detail, since those who may have been more baffled by the highly ritualistic sequences may have simply waffled more, or invented more false detail. We note that the study presented in Supplementary Material A (a precursor to the primary study presented herein) that word count positively correlated with memory for the sequence ($r = .305$), success ($r = .476$), and accuracy ($r = .728$). Importantly, accuracy was measured using a ~28 point criteria, and applied to free recall data - thus, accuracy is a proxy measure for detail. Thus, measures presented in Supplementary Material A show that word count - while not perfect - is a useful measure for detail in the present experimental context.

The null hypothesis was that ritualistic actions are neither costly nor beneficial. That is, as the proportion of rituals varied within a sequence, we would observe no effect on recall. However, sequences that featured high proportions of ritual were less accurately recalled than sequences with low amounts, and sequence-2 was many times more likely to be reported successfully than sequence-6. While the binary measures of 'success' appears less sensitive than 'accuracy', this may be due to the highly conservative nature of this measure: any single mistake in the ordering of action elements constituted a failure (acting as an analogue for a task that has no intermediate level of success, e.g., baking a soufflé, tying a bow-tie). This measure, as intended, may have a degree of ecological validity, but it suppresses a great deal of variance. That said, the highest rates of ritualized gesture harmed success (and low rates hinted at improvement). Participants reported decreasing confidence in their memory for the test-sequence as the proportion of rituals increased (and were remarkably conserved between the study presented in the supplementary material and the experiment proper; betas of -.172 and -.115, respectively). In other words, in

the primary study, for every additional ritualized gestures in the sequence, confidence declined by 11.5%. All in all, we have evidence to reasonably reject the null hypothesis, and can claim that rituals incur cognitive costs.

These findings need to be discussed in the context of both competing hypotheses. The 'catalyst' hypothesis predicts that, while ritualized elements are generally regarded as cognitively expensive, they offset their cost by providing a memory advantage for the sequence in which they are embedded. This benefit, however, is modest, such that after a certain threshold, the cost exceeds the benefit, and the presence of rituals should be harmful to performance. The alternative 'cost' hypothesis predicts that ritualistic action is cognitively expensive and is not offset by any benefit, thus, performance on recall measures should be (increasingly) diminished at all levels of ritualistic action. No analysis found an improvement for recall at low proportions above baseline, and yet participants in sequence-2 were over 3 times more likely to be successful than those in sequence-6. Thus, while we cannot argue that we have convincing evidence in favour of the 'catalyst' hypotheses, we do have a pattern of data broadly inconsistent with the cost hypothesis (as well as no evidence for any cost at the lowest proportions). And so, while we cannot with confidence endorse either alternative hypothesis, we find ourselves erring in favour of the catalyst hypothesis.

The question we must address then is: *how 'costly' are rituals, anyway*? It seems that the inclusion of ritualistic gesture incurs a cognitive cost, particularly at high levels. This is supported by participants' self-reported confidence in their performance as the proportion of ritualistic gestures increases, a finding we argue is evidence of the experience of cognitive load (Brunken, Plass, & Leutner, 2003). It seems likely, then, that low levels of ritualistic action generate load, but at levels too trivial to be important (alternatively, at levels too trivial to be detected by the present design and/or analysis). To the best of our collective knowledge, no prior attempts have been made to quantify how much cognitive load is generated by ritualized gesture (but for a computational model of this question, see: Kristoffer Nielbo & Sørensen, 2015; K. Nielbo & Sørensen, 2011). Thus, individual ritualized gestures are likely not particularly costly, but those costs are cumulative.

A question secondary to the theory, but relevant for interpretation is, h*ow good are people at remembering actions they observe?* As discussed earlier, the literature on overimitation reveals that both children and adults (Keupp et al., 2015; Lyons, Young, & Keil, 2007; McGuigan et al., 2011; Nielsen et al., 2014; Nielsen & Tomaselli, 2010) are very good at performing a series of actions after having it modelled to them for only a brief period. So good, in fact, that ceiling effects are a methodological issue (Keupp et al., 2015). However, the tasks used in overimitations paradigms are often quite simple (consisting of tapping, probing, and swiping which are not hierarchically nested), and (though infrequently reported) usually take between 30 - 60 seconds to model. The literature on Delayed Motor Intention (DMI) also notes that people are better at recalling an observed sequence when they have an expectation to *perform* rather than simply *recall* motor actions (Badets & Blandin, 2011; Badets et al., 2006; Badets & Osiurak, 2015). Indeed, participants who are given a written list of motor commands (e.g., pick up the rock, slide the pen) are more accurate when performing them than verbally recalling them (Koriat et al., 1990). Thus, it seems that motor action is privileged between (written -> enacted) and within (enacted ->

enacted) domains of recall. That our experiment presented participants with stimuli that was longer than usual (2 minutes), more complicated than is typical, and via a mode of recall independent of performance, imply two issues: first, that any benefit associated with ritualized actions may be understated, and second, that decline at high levels may be overstated. Thus, we must keep this in mind when understanding these findings - and hope that future research using behavioral paradigms can provide better resolution on the question.

And so what are the potential implications of cognitively costly ritualistic gestures? Consider Ritual Modes Theory (Whitehouse, 2004). 'Doctrinal rituals' are low arousal, high frequency events, in which participants routinely execute formal rituals in orthodox contexts, the purpose of which is to inculcate complex dogma. Whitehouse (2004) argues that frequent repetition in ritualised contexts aids semantic memory for these religious teachings. Further, that repetition of the ritual and dogma can lead to 'tedium' (Whitehouse, 2004; Whitehouse, Kahn, Hochberg, & Bryson, 2012) and uncritical practice of the act. Indeed, Whitehouse (2004) writes "*[Implicit] memory for religious rituals enhances the survival potential of authoritative teachings stored in semantic memory"* (p. 68). If so, according to the 'catalyst hypothesis', by including an increasing (but limited) number of rituals in doctrinal contexts, the uptake and retention (and subsequent) transmission of dogma may be enhanced. Which, in turn, may partly explain how doctrinal modes limit religious 'innovation'. However, if it is the case that any number of rituals is damaging to subsequent memories (consistent with the cost hypothesis), then the doctrinal mode may be successful despite its ritualisation[4], which in turn, may explain the presence of orthodoxy checks in centralized religions. Further, it may even be possible that high rates of rituals, with their additional cognitive load, are [one of] the mechanism(s) that lead to tedium and uncritical practice. In any event, whether the frequency of rituals in doctrinal contexts influences memory for dogma positively or negatively is an empirical question.

Along similar lines, Rossano (2012) argues that ritualized gestures, by virtue of their general uninterpretable - but intentional - nature are evolved cues that the associated normative/social information is relevant and important. A theoretical position consistent with the hypothesis of a memory advantage for semantic information at some, if not all, levels of ritual frequency. (Though likely only at low levels, since rituals act as an attention-cueing device, rather than a memory-related tool; Rossano, 2012). On the other hand, Schjoedt et al., (2013) appear to suggest that cognitive depletion, introduced by ritualised is a necessary condition for meaning making. More ritualized gestures are better, as they creates a larger 'inferential gap' for the ritual participant to 'fill' with orthodoxy thereafter (an argument that appears to apply for both imagistic and doctrinal modes). If they are correct, then Schjoedt et al., (2013) would predict in doctrinal contexts that there should be moderate to high amounts of ritual, which serve the goal of depletion (which will require the individual to fill with orthodoxy at a later point).

While multiple theories of ritual and cultural cognition specify a role for ritualized gesture, cognition, and memory, there is relatively little quantification of the phenomenon. Despite the fact that it appears to be a

---

[4] *With the caveat that there are many other important features of the Doctrinal Mode that can influence memory for dogma, see: Whitehouse (2004).*

relevant, even necessary, dimension of multiple theories. Here we are show that ritualized gesture at low proportions may be only nominally costly for social learners, though it has a cumulative effect, such that it appears to damage recall for the sequence overall at higher proportions. Nonetheless, we cannot reject the possibility - and may even argue that the evidence is suggestive - that low levels of ritualized gesture produce a benefit relative to its absence. We can, however, reject the null hypothesis, that rituals are free. We must conclude (rather typically) that more research on this topic is required. Though hopefully the implications of the present data are clear - in both modern and historical contexts - the inclusion of *too much* ritualisation is likely to impair learning (at least initially), which is likely to affect the transmission of important action sequences, and influence the shape of variation thereafter.

Acknowledgements

Reference list

Acerbi, A., & Tennie, C. (2016). The role of redundant information in cultural transmission and cultural stabilization. *Journal of Comparative Psychology*, *130*(1), 62–70.

Atkisson, C., O'Brien, M. J., & Mesoudi, A. (2012). Adult Learners in a Novel Environment Use Prestige-Biased Social Learning. *Evolutionary Psychology: An International Journal of Evolutionary Approaches to Psychology and Behavior*, *10*(3), 147470491201000.

Badets, A., & Blandin, Y. (2011). Feedback and intention during motor-skill learning: a connection with prospective memory. *Psychological Research*, *76*(5), 601–610.

Badets, A., Blandin, Y., & Shea, C. H. (2006). Intention in motor learning through observation. *The Quarterly Journal of Experimental Psychology*, *59*(2), 377–386.

Badets, A., & Osiurak, F. (2015). A goal-based mechanism for delayed motor intention: considerations from motor skills, tool use and action memory. *Psychological Research*, *79*(3), 345–360.

Bailey, H. R., Kurby, C. A., Giovannetti, T., & Zacks, J. (2013). Action perception predicts action performance. *Neuropsychologia*, *51*(11), 2294–2304.

Barrett, J., & Thomas Lawson, E. (2001). Ritual Intuitions: Cognitive Contributions to Judgments of Ritual Efficacy. *Journal of Cognition and Culture*, *1*(2), 183–201.

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *53*(1), 153–164.

Berl, R., & Hewlett, B. (2015). Cultural Variation in the Use of Overimitation by the Aka and Ngandu of the Congo Basin. *PloS One*, *10*(3). https://doi.org/e0120180.

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology. Applied*, *5*(1), 76–88.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*(12), 539–546.

Boyer, P., & Liénard, P. (2006). Why ritualized behavior? Precaution Systems and action parsing in developmental, pathological and cultural rituals. *The Behavioral and Brain Sciences*, *29*(6), 595–613; discussion 613–650.

Brennessel, B., Drout, M. D. C., & Gravel, R. (2005). A reassessment of the efficacy of Anglo-Saxon medicine. *Anglo-Saxon England*, *34*(-1), 183.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 53–61.

Carter, C. S., Botvinick, M. M., & Cohen, J. D. (1999). The contribution of the anterior cingulate cortex to executive processes in cognition. *Reviews in the Neurosciences*, *10*(1), 49–57.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press.

Cohen, M. D., & Bacdayan, P. (1994). Organizational Routines Are Stored as Procedural Memory: Evidence from a Laboratory Study. *Organization Science*, *5*(4), 554–568.

Csibra, G. (1993). Action mirroring and action understanding: an alternative account. In *Sensorimotor Foundations of Higher Cognition* (pp. 435–459).

Dennett, D. C. (1996). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Penguin UK.

Durkheim, É. (2012). *The Elementary Forms of the Religious Life*. Courier Corporation.

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, *11*(1), 19–23.

Flores, S., Bailey, H., Eisenberg, M. L., & Zacks, J. (2014). Effective Event Segmentation Improves Memory for Everyday Events Over Time. *PsycEXTRA Dataset*. https://doi.org/10.1037/e528942014-930

Flores, S., Bailey, H. R., Eisenberg, M. L., & Zacks, J. (2017). Event segmentation improves event memory up to one month later. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(8), 1183–1202.

George, D., & Mallery, P. (2011). *SPSS for Windows Step by Step: A Simple Guide and Reference 18.0 Update*. Pearson College Division.

Gigerenzer, G., & Selten, R. (2002). *Bounded Rationality: The Adaptive Toolbox*. MIT Press.

Gleissner, B., Bekkering, H., & Meltzoff, A. N. (2000). Children's coding of human action: cognitive factors influencing imitation in 3-year-old. *Developmental Science*, *3*(4), 405–414.

Gold, D. A., Zacks, J., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive Research: Principles and Implications*, *2*(1), 1.

Hamilton, A. F. de C., de C. Hamilton, A. F., & Grafton, S. T. (1993). The motor hierarchy: from kinematics to goals and intentions. In *Sensorimotor Foundations of Higher Cognition* (pp. 381–407).

Harrison, F., Roberts, A. E. L., Gabrilska, R., Rumbaugh, K. P., Lee, C., & Diggle, S. P. (2015). A 1,000-Year-Old Antimicrobial Remedy with Antistaphylococcal Activity. *mBio*, *6*(4), e01129.

Henrich, J. (2009). The evolution of costly displays, cooperation and religion. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *30*(4), 244–260.

Hewlett, B. S., Fouts, H. N., Boyette, A. H., & Hewlett, B. L. (2011). Social Learning Among Congo Basin Hunter–Gatherers*. In *Culture Evolves* (pp. 411–430).

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition*, *8*(3), 164–181.

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(5), 580–593.

Kapitány, R., & Nielsen, M. (2015). Adopting the ritual stance: The role of opacity and context in ritual and everyday actions. *Cognition*, *145*, 13–29.

Kapitány, R., & Nielsen, M. (2016). The ritual stance and the precaution system: the role of goal-demotion and opacity in ritual and everyday actions. *Religion, Brain & Behavior*, *7*(1), 27–42.

Kempe, M., & Mesoudi, A. (2014). An experimental demonstration of the effect of group size on cultural accumulation. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *35*(4), 285–290.

Keren, H., Boyer, P., Mort, J., & Eilam, D. (2013). The impact of precaution and practice on the performance of a risky motor task. *Behavioral Sciences*, *3*(3), 316–329.

Keupp, S., Behne, T., Zachow, J., Kasbohm, A., & Rakoczy, H. (2015). Over-imitation is not automatic: context sensitivity in children's overimitation and action interpretation of causally irrelevant actions. *Journal of Experimental Child Psychology*, *130*, 163–175.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166.

Kline, M. A. (2014). How to learn about teaching: An evolutionary framework for the study of teaching behavior in

humans and other animals. *The Behavioral and Brain Sciences*, 1–70.

Koriat, A., Ben-Zur, H., & Nussbaum, A. (1990). Encoding information for future action: memory for to-be-performed tasks versus memory for to-be-recalled tasks. *Memory & Cognition*, *18*(6), 568–578.

Legare, C. H., & Nielsen, M. (2015). Imitation and Innovation: The Dual Engines of Cultural Learning. *Trends in Cognitive Sciences*, *19*(11), 688–699.

Legare, C. H., & Souza, A. L. (2012). Evaluating ritual efficacy: evidence from the supernatural. *Cognition*, *124*(1), 1–15.

Legare, C. H., Wen, N. J., Herrmann, P. A., & Whitehouse, H. (2015). Imitative flexibility and the development of cultural learning. *Cognition*, *142*, 351–361.

Liénard, P., & Boyer, P. (2006). Whence Collective Rituals? A Cultural Selection Model of Ritualized Behavior. *American Anthropologist*, *108*(4), 814–827.

Lozada, M., Ladio, A., & Weigandt, M. (2006). Cultural Transmission of Ethnobotanical Knowledge in a Rural Community of Northwestern Patagonia, Argentina. *Economic Botany*, *60*(4), 374–385.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19751–19756.

Malle, B. F., Moses, L. J., & Baldwin, D. A. (2001). *Intentions and Intentionality: Foundations of Social Cognition*. MIT Press.

McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology*, *102*(1), 1–18.

Mesoudi, A. (2008). An experimental simulation of the "copy-successful-individuals" cultural learning strategy: adaptive landscapes, producer–scrounger dynamics, and informational access costs. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *29*(5), 350–363.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38.

Nielbo, K., Schjoedt, U., & Sørensen, J. (2012). Hierarchical Organization of Segmentation in Non-Functional Action Sequences. *Journal for the Cognitive Science of Religion*, *1*(1). https://doi.org/10.1558/jcsr.v1i1.71

Nielbo, K., & Sørensen, J. (2011). Spontaneous processing of functional and non-functional action sequences. *Religion, Brain & Behavior*, *1*(1), 18–30.

Nielbo, K., & Sørensen, J. (2015). Attentional resource allocation and cultural modulation in a computational model of ritualized behavior. *Religion, Brain & Behavior*, *6*(4), 318–335.

Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental Psychology*, *47*(4), 1050–1053.

Nielsen, M., Kapitány, R., & Elkins, R. (2015). The perpetuation of ritualistic actions as revealed by young children's transmission of normative behavior. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *36*(3), 191–198.

Nielsen, M., Mushin, I., Tomaselli, K., & Whiten, A. (2014). Where culture takes hold: "overimitation" and its flexible deployment in Western, Aboriginal, and Bushmen children. *Child Development*, *85*(6), 2169–2184.

Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science*, *21*(5), 729–736.

Nielsen, M., Tomaselli, K., & Kapitány, R. (2018). The influence of goal demotion on children's reproduction of ritual behavior. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*. https://doi.org/10.1016/j.evolhumbehav.2018.02.006

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: a meta-analysis. *Psychology and Aging*, *23*(1), 104–118.

Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, *126*(2), 182–192.

Palmer, C., & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science*, *11*(1), 63–68.

Radvansky, G. A. (2012). Across the Event Horizon. *Current Directions in Psychological Science*, *21*(4), 269–272.

Richert, R., Whitehouse, H., & Stewert, E. (2005). Memory and analogical thinking in high-arousal rituals. In *Mind and religion: Psychological and cognitive foundations of religiosity* (pp. 127–145).

Rosenbaum, D. A., & Gregory, R. W. (2002). Development of a method for measuring movement-related effort: biomechanical considerations and implications for Fitts' law. *Experimental Brain Research. Experimentelle*

Hirnforschung. *Experimentation Cerebrale*, *142*(3), 365–373.

Rossano, M. J. (2012). The essential role of ritual in the transmission and reinforcement of social norms. *Psychological Bulletin*, *138*(3), 529–549.

Ruffle, B. J., & Sosis, R. (2007). Does It Pay To Pray? Costly Ritual and Cooperation. *The B.E. Journal of Economic Analysis & Policy*, *7*(1). https://doi.org/10.2202/1935-1682.1629

Ruffle, B. J., & Sosis, R. H. (2003). Does it Pay to Pray? Evaluating the Economic Return to Religious Ritual. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.441285

Rybanska, V., McKay, R., Jong, J., & Whitehouse, H. (2017). Rituals Improve Children's Ability to Delay Gratification. *Child Development*. https://doi.org/10.1111/cdev.12762

Sargent, J. Q., Zacks, J., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., … Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255.

Schjoedt, U., Sørensen, J., Nielbo, K., Xygalatas, D., Mitkidis, P., & Bulbulia, J. (2013). Cognitive resource depletion in religious interactions. *Religion, Brain & Behavior*, *3*(1), 39–55.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225–260.

Shea, C. H., & Wulf, G. (2005). Schema theory: a critical appraisal and reevaluation. *Journal of Motor Behavior*, *37*(2), 85–101.

Sosis, R., & Bressler, E. R. (2003). Cooperation and Commune Longevity: A Test of the Costly Signaling Theory of Religion. *Cross-Cultural Research: Official Journal of the Society for Cross-Cultural Research / Sponsored by the Human Relations Area Files, Inc*, *37*(2), 211–239.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging*, *10*(4), 527–539.

Swallow, K. M., Zacks, J., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology. General*, *138*(2), 236–257.

Tennie, C., Call, J., & Tomasello, M. (2006). Push or Pull: Imitation vs. Emulation in Great Apes and Human Children. *Ethology: Formerly Zeitschrift Fur Tierpsychologie*, *112*(12), 1159–1169.

van Mulukom, V. (2017). Remembering religious rituals: autobiographical memories of high-arousal religious rituals considered from a narrative processing perspective. *Religion, Brain & Behavior*, *7*(3), 191–205.

Voth, C. (2017). Bald's Leechbook. In *The Encyclopedia of Medieval Literature in Britain* (pp. 1–3).

Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007). Gender differences in object location memory: a meta-analysis. *Psychonomic Bulletin & Review*, *14*(1), 23–38.

Watson-Jones, R. E., Legare, C. H., Whitehouse, H., & Clegg, J. M. (2014). Task-specific effects of ostracism on imitative fidelity in early childhood. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *35*(3), 204–210.

Weber, B. H., & Depew, D. J. (2003). *Evolution and Learning: The Baldwin Effect Reconsidered*. MIT Press.

Wen, N. J., Herrmann, P. A., & Legare, C. H. (2016). Ritual increases children's affiliation with in-group members. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *37*(1), 54–60.

Whitehouse, H. (2004). *Modes of Religiosity: A Cognitive Theory of Religious Transmission*. Rowman Altamira.

Whitehouse, H., Kahn, K., Hochberg, M. E., & Bryson, J. J. (2012). The role for simulations in theory construction for the social sciences: case studies concerning Divergent Modes of Religiosity. *Religion, Brain & Behavior*, *2*(3), 182–201.

Whiten, A., Allan, G., Devlin, S., Kseib, N., Raw, N., & McGuigan, N. (2016). Social Learning in the Real-World: "Over-Imitation" Occurs in Both Children and Adults Unaware of Participation in an Experiment and Independently of Social Interaction. *PloS One*, *11*(7), e0159920.

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1528), 2417–2428.

Wilder, D. A. (1978a). Effect of Predictability on Units of Perception and Attribution. *Personality & Social Psychology Bulletin*, *4*(2), 281–284.

Wilder, D. A. (1978b). Predictability of Behaviors, Goals, and Unit of Perception. *Personality & Social Psychology Bulletin*, *4*(4), 604–607.

Wood, L. A., Harrison, R. A., Lucas, A. J., McGuigan, N., Burdett, E. R. R., & Whiten, A. (2016). "Model age-based" and "copy when uncertain" biases in children's social learning of a novel task. *Journal of Experimental Child Psychology*, *150*, 272–284.

Wood, L. A., Kendal, R. L., & Flynn, E. G. (2012). Context-dependent model-based biases in cultural transmission:

children's imitation is affected by model age over model knowledge state. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *33*(4), 387–394.

Zacks, J. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979–1008.

Zacks, J., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293.

Zacks, J., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging*, *21*(3), 466–482.

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.

Zacks, J., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology. General*, *130*(1), 29–58.

Supplementary Material A.

S1.1 Introduction

The following study was conducted prior to the study presented in the manuscript proper. Due to aspects of the design, and the consequent limitations on interpretation, our peer-reviewers suggested we move this study into the supplementary material. What follows is a description of the pre-registered methods, results, and analyses of our initial experiment. Only minor changes, primarily methodological, were introduced into the study presented in the body of the manuscript.

S1.2 Predictions.

1. We predict that some small ratio of redundant actions (e.g, Sequence-1, -2, and perhaps -3) will significantly improve Accuracy and Success for the whole sequence above baseline (sequence-0).
2. We predict that a high ratio of redundant actions (e.g., Sequence-6, -5, and perhaps -4) will result in lower Accuracy and Success for the whole sequence compared to baseline (Sequence-0).
3. We predict that sequences with small amounts of redundancy will have increasing amount of Detail compared to baseline. *i.e.,* there will be more detail in Sequence-1 than in Sequence-0, and more in Sequence-2 than Sequence-1, etc.
4. We anticipate that this is a curvilinear relationship, and that once cognitive load reaches an unspecified threshold, detail will decline.

S2.1 Methods.

S2.2 Stimuli.

Two complex 120-second recipe-like action sequences were created, each of which consisted of six sub-goals that needed to be satisfied in order to produce a final product. Two 'master sequences' were created, which exclusively featured instrumental action (there were no ritualistic features; 'Sequence-0'). Six variations were made of each master-sequence, where increasing proportions of the sub-goals were made ritualistic. The first variation (Sequence-1) featured 1 ritualistic action and 5 instrumental actions, the second variation (Sequence-2) features 2 ritualistic actions, and 4 instrumental actions, and so on. The sixth and final sequence (Sequence-6) featured only ritualistic actions. In total, there were 7 variations for each master sequence. We provided participants with the cover-story that both sequences were medieval recipes, where one sequence produced a drink for the treatment of *worms* (a gut parasite), and the other sequence produced a paste for the treatment of a *burn*. Table 1 (of the manuscript proper) describes each sequence, and what the instrumental and ritualistic actions were in each. It is worth noting that when a ritualistic act is first presented (in, say, sequence-1), that ritualistic action is preserved in all subsequent variations (in sequence-2, -3, -4, etc.). Put another way, this experiment did not control for *where* in

the sequence a ritualistic action was performed, but manipulated the *frequency* of ritualistic action in the sequences overall. Links to all videos are in Supplementary Material B.

S2.3 Pre-Registration.

This experiment was pre-registered as part of the *'$1,000 Challenge'* on the Open Science Framework. The registration file is available at https://osf.io/dhy76/registrations/. Here we describe the pre-registered data collection plan, analysis plan, and predictions.

S2.4 Data Collection and procedure.

We recruited 350 participants on the crowd-sourcing platform 'Prolific' (similar to mTurk). Participants were paid £1.30 (GBP) for their participation in the 12 - 15 minute study. Participants were eligible if they lived in the UK or the US, were native English speakers, were older than 18, and used a PC to complete the experiment (not a tablet or smaller device). Participants watched one random video from each of the two sequences (hereafter, *'worms'* and *'burns'*) twice. Thereafter, participants were provided with a still image of the initial layout of the table and ingredients and were asked to describe what they saw '*In as much detail as possible... The goal here is to write a description that is sufficiently detailed so that someone who had not seen the video could recreate the sequence accurately.*'

S2.5 Participants.

We initially recruited 350 participants, anticipating to have 50 participants per condition. This number was dictated by available financial resources and the length of the survey (rather than a more traditional power analysis). Of the original 350, a total of 19 participants were excluded from analysis as they did not complete any of the key measures. The final analysis included data from 331 individuals, where the mean age was 32.39 years (SD = 11.72), in which 161 were women, 166 were men, and 4 self-described as 'other'[5].

S2.6 Dependent variables and Coding.

Proportional Accuracy. As shown in supplementary Table 1 and 2 of Supplementary Material E, if a participant described an action that did occur, the action was awarded a single point (unlike the 'success' measure', this includes descriptions of ritualistic actions). A maximum score for each sequence was calculated. While an *a priori* attempt was made to ensure the same maxima for the *worms* and *burns* sequences, this was not possible. Scores for each sequence may vary by a small degree (See Supplementary Material E). To control for this, these absolute scores were converted into proportions to allow for analysis. This measure was not sensitive to ordering errors, but was a measure of whether or not the participant described given action-elements as present

---

[5] We observed that the low N violated many assumptions of analyses, and for the sake of robust analysis, 'Other' has been treated as missing data.

or not.

Binary Success. Based on the coding scheme described supplementary Table 1 and 2 of Supplementary Material E, an evaluation was made whether the participant had described the necessary causal actions, and the correct quantities or durations, in sufficient detail so as to produce the desired outcome. This measure (unlike '*proportional accuracy*', described below) is a proxy for determining not only whether elements were recalled accurately, but that the were nested in the correct manner in order to produce a successful outcome - thus making this a conservative measure that reflects some learning situations in which one cannot partially succeed (e.g., one cannot partially succeed in baking a soufflé, one succeeds or one fails). Ritualized elements were not coded here as they were not instrumentally relevant.

The coder followed the instructions: Any omission errors (e.g., forgetting a step or ingredient), or underspecified quantity errors (e.g., two pinches of tobacco rather than three), resulted in a fail score. Hierarchical order errors also resulted in a fail score (e.g., if a participant incorrectly described adding an ingredient *after* heating the mixture rather than *before*, then a fail score was assigned. However, if a participant described a whole subgoal out of order, but it didn't influence the hierarchical contingency of earlier/later steps, this was acceptable). Additionally, though infrequent, any mislabeled ingredients resulted in a fail score (e.g., describing the *Tobacco* as *Rosemary*). As shown in supplementary Table 1 and 2, each action was described as necessary or not (as ritualistic actions are not, by definition, causally necessary). To achieve a 'pass score' all instrumental actions needed to be described in appropriate detail. Thus, 'Binary Success' is a highly conservative measure.

Detail. This was a simple word count (but see: Pre-registration). Notably, we did not follow through with our pre-registered approach for 'Detail'. We had intended a very rich extraction of 'detail', however it became apparent that this scheme was too complicated and time-consuming to be successfully executed. We opted, instead, to simply take a word count as a coarse measure of expressed detail.

S3.1 Results

S3.2 Hypothesis 1 & 2: Accuracy (Pre-Registered).

We specified the use of a between-subjects ANOVA on our measure of Accuracy. As prior research has shown that women tend to have better memory than men (Voyer, Postma, Brake, & Imperato-McGinley, 2007), and that age may negatively influence memory (Old & Naveh-Benjamin, 2008; Spencer & Raz, 1995), we controlled for participant age and sex (thus, we executed an ANCOVA rather than an ANOVA). We entered in condition information, as well as which sequence the participant observed. We combined responses from the worms and burns sequences. We did not find a main effect of condition, $F(6, 625) = 1.077$, $p = .375$ (See Figure S1 for distributions of scores across conditions). We observed a main effect of gender, $F(1, 625) = 18.257$, $p < .001$, such that women ($M = .589$, $SD = .203$) performed better than men ($M = .514$, $SD = .220$), $d = .354$, *95%CI* -0.107 - -0.040. We did not find an effect of age, $F(1, 625) = .086$, $p = .770$. We did observe a main effect of which sequence participants observed, $F(1, 625) = 4.818$, $p = .029$, such that the Worms sequence ($M = .572$, $SD = .213$) was better

recalled than the Burns sequence (*M* = .535, *SD* = .217), *d* = .172, *95%CI* .004 - .070. The same analyses were run on worms and burns sequences individually, and the same pattern of results was observed (see Supplementary Material D). These results did not provide support for Hypothesis 1 and 2 - as the number of ritualized gestures embedded within instrumental sequences did not improve the proportional accuracy in accurately recalling sequence elements at low frequencies, nor did it negatively impact accuracy in recalling sequence elements at high frequencies.

Figure S1. *Descriptive statistics for Accuracy across Conditions.*
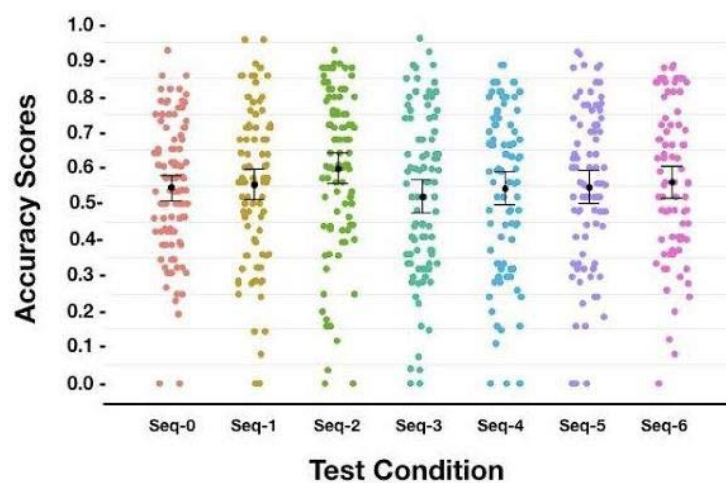
S3.2 Hypothesis 1 & 2: Success (Pre-registered).

We specified the use of a between-subjects ANOVA on our measure of Success. However, 'success' was a binary variable, and this violates the assumptions of the test. Rather, we conducted the more appropriate Multinomial Logistic Regression. We entered in condition information, and controlled for participant age and gender, as well as which sequence the participant observed. We combined responses from the worms and burns sequences. We did not find a significant goodness-of-fit values, $\chi^2(9)$ = 11.313, *p* = .255. Figure S2 shows the proportion and absolute number of 'successful' descriptions of the each sequence observed.

The results of these analyses are not in support of Hypothesis 1 and 2. The number of ritualized gestures embedded within overall sequences neither improved successful retelling at low frequencies, nor did it negatively impact successful retelling at high frequencies.

Figure 2. *The proportions of 'successful' descriptions of the each sequence observed.*
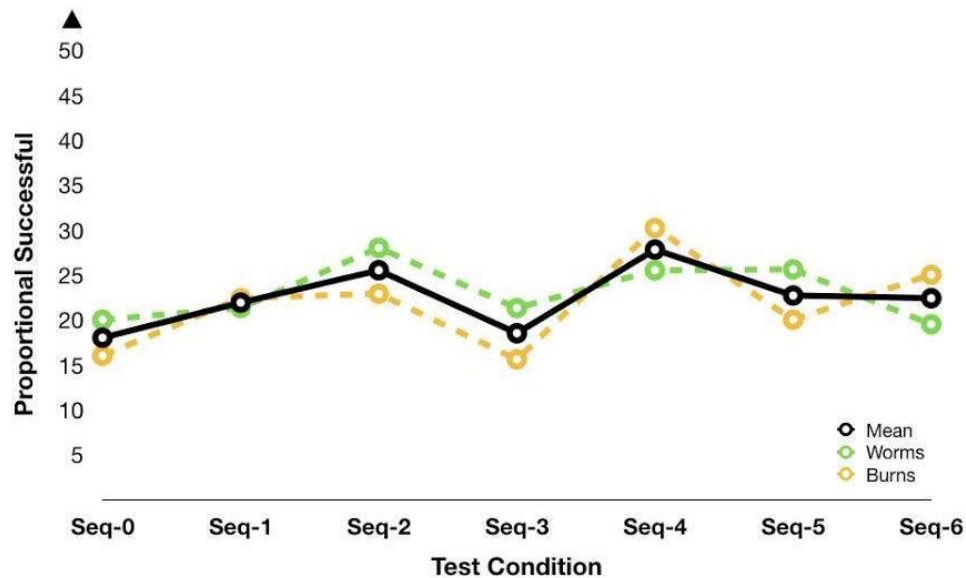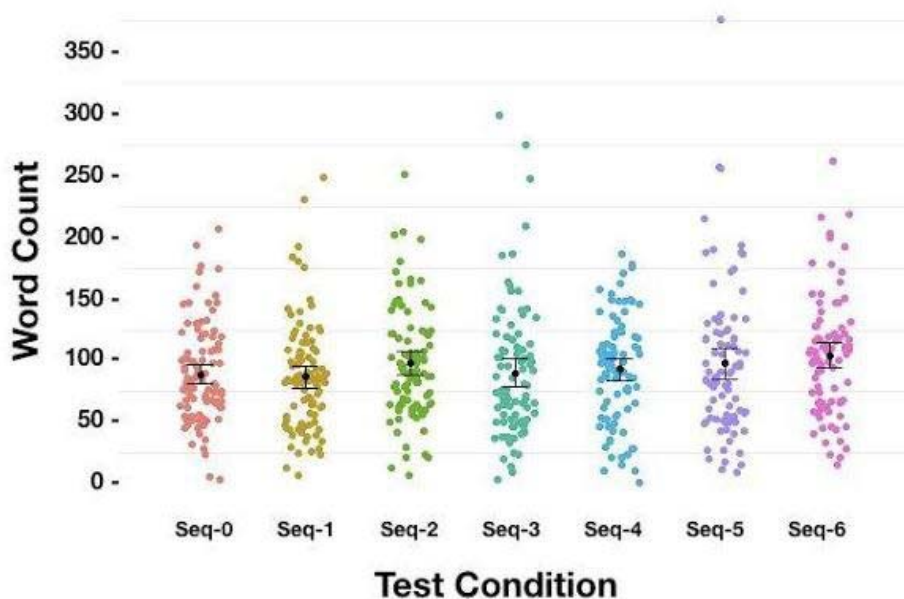

S3.3 Hypothesis 3 & 4: Detail (Pre-registered).

Hypothesis 3: We predict that sequences with small amounts of redundancy will have increasing amount of Detail compared to baseline. *i.e.,* there will be more detail in video 1 than in video 0, and more in video 2 than video 1.

Hypothesis 4: We anticipate that this is a curvilinear relationship, and that once cognitive load reaches an unspecified threshold, detail will decline.

We specified the use of a between-subjects ANOVA on our measure of detail. We entered in condition information, and controlled for participant age and sex, as well as which sequence the participant observed. We did not find an effect of condition, $F(6, 625) = 1.476$, $p = .184$ but there was an effect of sex, $F(1, 325) = 11.780$, $p = .001$, such that women ($M = 99.99$, $SD = 49.09$) wrote more words than men ($M = 56.65$ $SD = 45.60$). There was no effect of age, $F(1, 625) = .283$, $p = .595$. There was also an effect of sequences, $F(1, 625) = 30.421$, $p < .001$, such that participants wrote more words for the worms sequence ($M = 104.12$, $SD = 52.30$) than the burns sequences ($M = 83.78$, $SD = 42.48$). Figure S3 shows the distribution of word count scores (while Supplementary Material D Table 3 shows the means and standard deviations of each sequence for each video).

Figure S3. *Distribution of word counts across conditions collapsed across master sequences.*

S3.4 Hypothesis 4: Detail (Exploratory).

Our pre-registered hypothesis predicted a decline in detail at high ratios of ritual elements. This was predicted as we anticipated that the cognitive load associated with high ratios would swamp participants' ability to describe the sequence. However, the pattern in the data was to the contrary, thus, we conducted post-hoc follow-up analysis. On data collapsed across the worms/burns sequence, contrasts revealed that Sequence-6 was significantly greater than Sequence-0, $p = .018$, *95%CI* 3.558 - 37.466, and significantly greater than Sequence-1, $p = .036$, *95%CI* 1.228 - 35.465. No other differences were observed.

S3.5 Descriptive and Exploratory Analysis.

Participants were asked three questions after providing responses on the primary dependent variables. Those questions were: "*How good is your memory generally?*" and "*How good was your memory for the [Burns]/[Worms] recipe?*". This was measured on a 100-point scale, ranging from 0 (Terrible) to 100 (Excellent) with a midpoint of 50 (Average). We conducted a linear regression to examine whether there condition assignment predicted participants' self-reports of their general memory (*M* = 61.16, *SD* = 19.84), and we controlled for age and sex. The model produced did not fit, $F_{(3, 315)} = .195$, $p = .900$, suggesting that the general measure was not influenced by the experimental design.

We conducted a linear regression on participants' self-reported memory for the experimental sequence/condition participants were in, while controlling for age and sex. The model produced was significant, $F_{(3, 631)} = 6.520$, $p < .001$. Neither Age, $p = .401$, *95%CI* -0.211 - .085, nor Sex, $p = .647$, *95%CI* -2.666 - 4.287, provided significant contributions to the model. Condition, however, displayed a significant relationship $β = -0.172$,

p < .001, *95%CI* -2.774 - -1.055. The negative beta value indicates that as the number of rituals increased, self-reported memory for the specific sequence decreased. Table 2 shows the correlation matrix of dependent variables.

Table 2. *A correlation of all measured variables.*

			1	2	3	4	5
1	**Memory (General)**	correlation	-	.600**	.084*	.202**	.150**
2	**Memory (Condition)**	correlation		-	.241**	.406**	.305**
3	**Success (Binary)**	correlation			-	.607**	.476**
4	**Proportional Accuracy**	correlation				-	.728**
5	**Word count**	correlation					-

* p < .05, ** p < .001

S4.1 Discussion

We presented participants with two 2-minute videos of a novel action sequence, in which the goal was to produce either a drink (worms) or paste (burns). Each sequence had a varying proportion of ritualistic elements. The 'Catalyst hypothesis' predicted that action sequences with a small proportion of ritualistic features would be better (freely) recalled overall than sequences with no ritualistic elements on three measures: overall 'success', accuracy of recall on individual elements, and detail (word count). Moreover, we predicted that sequences with higher proportions of ritualistic action (sequence-4, -5, -6) would be lower on these measures overall than sequences with lower proportions (Sequence-0, -1, -2). Our results indicate that the proportion of ritualistic action embedded within a sequence did not appear to affect any of our measures.

It is possible that such null effects were observed because there is no true effect. However, it is also possible that no effect was observed because we lacked sufficient power to do so, due to elements of our design. Due to an oversight, we could not control for a participants' baseline ability to recall a given sequence. That is, we did not collect a universal baseline measure. A consequence of this is that our aggregated analyses also contain some non-independence (specifically, while condition was independent for each observation, age and sex were entered twice. The analyses in the supplementary material indicate this did not impact the results). Further, our dependent variables may not have allowed us to detect the true magnitude of the effect. Free recall for motor sequences is demonstrable less reliable than asking people to *perform* a motor sequence (Badets & Blandin, 2011; Badets et al., 2006; Badets & Osiurak, 2015). Thus, an additional pre-registered experimented is warranted, due to these oversights.

However, we did make several useful observations. First, participants reported that their confidence in

their own recall for a specific action sequence declined as the ratio of ritualistic gestures increased, indicating a subjective experience of difficulty in recalling what they observed. Second, word count positively correlates with memory for the sequence ($r = .305$), success ($r = .476$), and accuracy ($r = .728$). That is, those who write more are more confident and are justified in doing so, as they are also more successful, and more accurate (while those who write less appear to honestly report diminishing confidence, and are similarly justified).

We have refined our approach in a subsequent examining of the stated hypothesis, which are fully described in the manuscript proper, which were also pre-registered (osf.io/dhy76).

Supplementary Material B

Links to Stimuli

Sequence 1 0r6i https://youtu.be/ENRjtqHxlaA

Sequence 1 1r5i https://youtu.be/6lNs9xAib0A

Sequence 1 2r4i https://youtu.be/JSvgC6xEnaQ

Sequence 1 3r3i https://youtu.be/tA2Jz_i9c80

Sequence 1 4r2i https://youtu.be/Mu_KI-xn3uE

Sequence 1 5r1i https://youtu.be/6azDVctPydQ

Sequence 1 6r0i https://youtu.be/r7el7c1UNnQ

Sequence 2 0r6i https://youtu.be/CShOwQ5G3Jg

Sequence 2 1r5i https://youtu.be/R1A9WpZhhYs

Sequence 2 2r4i https://youtu.be/p-tyFn1LLyc

Sequence 2 3r3i https://youtu.be/mbcIeE53oQI

Sequence 2 4r2i https://youtu.be/G4nxHNQF6S8

Sequence 2 5r1i https://youtu.be/6GsclCCbUQ0

Sequence 2 6r0i https://youtu.be/klmbCTM8_AY

Supplementary Material C.

In study 2, participants were asked four questions on a 7-point scale (from strongly disagree to strongly agree) on the following statements: "The final product was physically ordinary, like the product of any other recipe", "The final product was special in a way other physical products are not", "I expect the final product to be an effective medicine for treating [worms/burns]", "The final product had a supernatural quality about it". These questions were asked in order to follow-up on results obtained by (Kapitány & Nielsen, 2015, 2016), who found that ritual actions did not change perceptions of ordinariness, but did appear to influence perceptions of 'specialness'. 'Supernaturalness' was measured as we predicted it to explain the effect of 'specialness' (if present). Finally, we predicted that perceived efficacy may be influenced, as previously (Vohs, Wang, Gino, & Norton, 2013) had found that rituals (rather than arbitrary actions) enhanced subjective qualities associated with food. supplementary Table 5 shows the correlations between these variables (in order to determine for collinearity). No collinearity was observed (Field, 2013), though supernaturalness and specialness were strongly related.

supplementary Table 5.

	Ordinary	Special	Supernatural	Efficacy
Ordinary	-	-.405**	-.349**	.069
Special	-	-	.610**	.411**
Supernatural	-	-	-	.161**
Efficacy	-	-	-	-

We conducted a linear regression on participants scores for whether the final product of the sequence was 'ordinary' (M = 4.41, SD = 1.92). We entered which condition they first observed, condition information, participant age and gender. We did not find a significant fit, $F(4, 402)$ = .369, p = .811.

Using the same set of predictors, as well as 'supernaturalness' we looked at perceptions of specialness (M = 3.66, SD = 1.93). We found a significant fit, $F(5, 400)$ = 50.082, $p < .001$. The effect of condition was significant, though only just, = -.079, t = 1.987, p = .048, *95%CI* -0.137 - -0.001 (indicated that the burns recipe was perceived as less supernatural than the worms recipe). As expected, perceptions of supernatural (M = 3.55, SD = 2.14) were significant, = .632, t = 15.568, $p < .001$, *95%CI* .499 - .644. No other effects were observed. We also

conducted the regression without supernaturalness as a predictor, an the model did not fit, $F(4, 402) = 1.269$, $p =$ .282.

We analysed perceptions of supernaturalness with sequence observed, condition, age and gender as predictors. The model was significant, $F(4, 402) = 7.240$, $p < .001$. We observed that there was an effect associated with condition, $= .182$, $t = 3.77$, $p < .001$, *95%CI* .083 - .265. We also observed an effect of which sequence participants observed, $= .159$, $t = 3.289$, $p = .001$, *95%CI* .273 - 1.084 (indicating that burns recipe led to more supernatural observations). Neither age nor gender predicted supernaturalness.

Finally, we examined perceptions of efficacy ($M = 3.57$, $SD = 1.86$). We entered in age and gender, which sequence the participant observed, condition information, as well as ordinariness, specialness, and supernaturalness. The model was significant, $F(7, 396) = 21.780$, $p < .001$. Neither age nor gender predicted outcomes. The sequence participants observed significantly predicted efficacy, $= -0.143$, $t = 3.300$ $p = .001$, *95%CI* -0.850 - -0.215 (indicating that burns recipe led to perceptions less efficacy). Condition, also, significantly predicted efficacy $= -0.133$, $t = 3.031$ $p = .003$, *95%CI* -0.182 - -0.039. As did ordinariness, $= 0.285$, $t = 5.990$ $p < .001$, *95%CI* .185 - .385. Surprisingly, specialness predicted efficacy, $= .560$, t = 9.924 $p < .001$, *95%CI* .430 - .643, but supernaturalness did not, $= -0.031$, $t = -0.540$ $p = .590$, *95%CI* -0.123 - 0.070.

Thus, in examining the collected data, and conducting analyses based on Kapitany and Nielsen (2015, 2016), we observed that the proportion of rituals did not influence how ordinary the final product was, and somewhat surprisingly, that the proportion of rituals did not predict how special an object was perceived to be. The proportion of rituals did predict specialness (marginally), but only if we accounted for perceptions of supernaturalness, which accounted for a lot. We measured perceived efficacy of the medicine, as it was consistent with the cover story, rather than preference, and found that ordinariness (positively predicted) and proportion of rituals (negatively predicted) suggesting that participants based perceptions of efficacy on causal intuitiveness and physical principles. And yet, perceptions of specialness (which was otherwise uninfluenced by proportion of ritual) strongly predicted efficacy, yet supernaturalness (which strongly predicted specialness) did not.

Supplementary Material D

Hypothesis 1 & 2: Accuracy on Separate Sequences

We conducted between-subjects ANOVA on our measure of Accuracy for the Worms sequence. We entered in condition information, and controlled for participant age and gender. We did not find a main effect of condition, $F(6, 309) = .669$, $p = .674$. We did observed a main effect of gender, $F(1, 309) = 5.521$, $p = .019$. Such that women ($M = .60$, $SD = .20$) performed better than men ($M = .54$, $SD = .22$), *95%CI* -0.104 - -0.009. We did not find an effect of age, $F(1, 309) = 1.204$, $p = .273$.

We conducted between-subjects ANOVA on our measure of Accuracy for the Burns sequence. We entered in condition information, and controlled for participant age and gender. We did not find a main effect of condition, $F(6, 308) = .357$, $p = .357$. We observed a main effect of gender, $F(1, 308) = 13.882$, $p < .001$. Such that women ($M = .58$, $SD = .20$) performed better than men ($M = .49$, $SD = .22$), *95%CI* -0.139 - -0.043. We did not find an effect of age, $F(1, 308) = .468$, $p = .494$.

Supplementary Table 3. *Means and Standard Deviation of Proportional Accuracy scores in retelling action elements across conditions.*

Proportional Accuracy			
Sequence	**Worms**	**Burns**	**Mean**
Sequence-0	.600 (.190)	.493 (.162)	.545 (.183)
Sequence-1	.561 (.201)	.548 (.227)	.554 (.214)
Sequence-2	.603 (.217)	.594 (.223)	.599 (.219)
Sequence-3	.546 (.249)	.496 (.212)	.523 (.232)
Sequence-4	.547 (.213)	.540 (.245)	.543 (.227)
Sequence-5	.562 (.212)	.534 (.231)	.548 (.221)
Sequence-6	.584 (.209)	.539 (.213)	.561 (.210)

Supplementary Table 4. *Means and Standard Deviations of Detail (wordcount) in recalling action sequences across conditions.*

Detail (word count)			
Sequence	**SEQ1**	**SEQ2**	**Total**
Sequence-0	103.06 (42.06)	74.42 (31.38)	88.64 (39.63)
Sequence-1	96.21 (43.40)	78.51 (43.92)	87.18 (44.34)
Sequence-2	112.70 (54.70)	82.29 (30.72)	97.81 (46.93)
Sequence-3	97.85 (57.64)	82.42 (50.38)	90.34 (54.47)
Sequence-4	93.17 (41.21)	92.91 (46.21)	93.04 (43.42)
Sequence-5	112.86 (69.50)	83.11 (47.65)	97.76 (60.86)
Sequence-6	114.48 (52.57)	95.05 (43.46)	104.42 (48.77)

Supplementary Material E.

(see here:

https://docs.google.com/document/d/1m1Sv33bAYUiumjl_I2Dvut5W8vOOBbs3JcG2NU4FEvs/edit?usp=sharing)