Variation matters: Expanding the scope of experimental archaeology using the Perception-Process-Product conceptual framework

Cheng Liu*

2024-02-05

Abstract

This paper presents the Perception-Process-Product ('Triple P') framework that aims to expand the scope of experimental archaeology. The Triple P framework emphasizes multi-level variation and interactions across the levels of perception, process, and product to provide a more grounded and richer explanation of the past archaeological record. It consists of three principles: 1) acknowledging the inherent trade-off between control and generalizability in the experimental research design; 2) encouraging collaborative projects that involve geographically diverse and non-traditional research participants such as hobbyists and novices; 3) adopting a workflow that normalizes the collection and curation of ethological and ethnographic data in experimental projects.

Keywords: Experimental archaeology; Ethological analysis; Ethnographic research; Curse of knowledge; Collaborative knowledge production

Contents

*Department of Anthropology, Emory University, Atlanta, GA, USA; cheng.liu@emory.edu

1 Introduction

This paper presents the Perception-Process-Product (hereafter 'Triple P') conceptual framework to expand the scope of experimental archaeology. The field has long tended to adopt the principle of Occam's razor (Blessing & Schmidt, 2021; e.g., Domínguez-Rodrigo, 2008; Reeves et al., 2009; P. Schmidt et al., 2019), whether explicitly or implicitly, and center around the reverse engineering of a past technology in a minimal or least-effort manner while ignoring the rich contextual information it affords. When applied to the experimental study of ancient craftsmanship, it implies that a technological solution that is simpler to reproduce is more likely to be the one used in the archaeological context. Nevertheless, Occam's razor, or the law of parsimony, can be insufficient to infer the preferences of 'irrational' agents possessing incomplete information (Mindermann & Armstrong, 2018) in tool design and use. The two conditions described here provide a better approximation of past humans displaying extensive cultural variation as opposed to the assumption of omniscient *Homo economicus* (i.e., the idea that humans are consistently rational and narrowly self-interested agents pursuing for optimality) for many anthropologists (Apicella et al., 2020; Henrich et al., 2001). Parallelly, Heyes (Heyes, 2012) also questioned the abuse of parsimony in animal behavioural research and further proposed that new observational and experimental studies that allow differential predictions to be tested become necessary when both a simple and a complex mechanism can explain the phenomenon of interest.

In the evolution of technology, it is rather common that opaque causal perception and its resulting tendency of over-imitation together with path dependence can lead to the widespread and long-lasting reproduction of technological solutions that are neither minimal in manufacture complexity nor optimal in functional efficiency. Over-imitation means the copying of actions that are causally irrelevant in a goal-directed action sequence (Lyons et al., 2007). It is a psychological propensity that was suggested to be uniquely prevalent among humans in inter-species (Clay & Tennie, 2018; Horner & Whiten, 2005) comparisons and cross-cultural contexts (Nielsen & Tomaselli, 2010; Stengelin et al., 2020). Gergely and Csisbra (2006) introduced 'Sylvia's Recipe' that vividly illustrates this cognitive process in the transmission of technical skills. Sylvia is an education researcher who developed a unique way of cooking ham roast by observing her mother during childhood, where she cut both ends of a ham. Later in life her mother happened to watch her cooking, where she noticed and questioned the purpose of this step of preparation. Sylvia could not answer it and was then told that it was processed that way because her mother did not have a pan that was large enough to cook a ham. The commonality of this opaque causal perception has also been demonstrated in a recent study of Hadza bowmakers. Harris et al. (2021) found that even experienced bowmakers only possess limited causal knowledge regarding the design and construction of bows according to modern engineering principles, meaning they cannot spell out the mechanical (dis)advantages of many morphological features.

On the other hand, path dependence also constrained the pursuit of functional optimization, where people are implicitly or explicitly aware of the existence of a more efficient solution but still stick to the older one due to the cost of learning, cultural conservatism (Acerbi et al., 2009; Ghirlanda et al., 2006; Morin, 2022), or other reasons. One such example in the evolution of technology is the longevity of QWERTY keyboard design (Kafaee et al., 2022). This deliberately unergonomic solution was invented in the era of typewriters in order to disperse commonly used letters, preventing the most frequently struck 'hammers' from clashing. Yet it is still the most common keyboard design today when such constraint does not exist anymore on modern computer hardware. In short, we should acknowledge the existence and variation of many "good-enough" technological solutions featuring various degrees of "redundancy" in real-world contexts, which often represent locally adaptive peaks instead of a global optimum in a multimodal fitness landscape due to multiple constraints and trade-off factors (Bettinger & Baumhoff, 1982; Mesoudi & O'Brien, 2008).

Built upon the *Homo economicus* critique as well as intellectual principles and practices in behavioural archaeology (Schiffer, 2010), here I propose the Triple P framework, which aims to **a)** amplify the expression of variation in experimental replicas (product) and their associated behavioural channels (process) as well as sensory experiences (perception) by experiments in diverse contexts and **b)** better identify the complex interacting relationships across these three levels of variations in real-world conditions. To accomplish these two objectives, I advocate the following three principles as integral components of the Triple P framework, which requires **1)** acknowledging the inherent trade-off between control and generalizability in the experimental research design and **2)** encouraging collaborative projects that involve geographically diverse and non-traditional research participants such as hobbyists and novices. These two principles are developed to advocate a pluralistic approach to the explanation of complex variation, which has received more attention from evolutionary anthropology (Antón & Kuzawa, 2017) to cognitive science (Barrett, 2020), instead of treating the optimization-based research agenda as a panacea.

The second principle particularly allows researchers to develop research questions that are also meaningful to descendant communities through respectful conversation and collaboration (Montgomery & Fryer, 2023). The Triple P framework also **3)** adopts a workflow that normalizes the collection and curation of ethological and ethnographic data in experimental projects. It is acknowledged that strategies of data collection and analysis of a given experimental project should be primarily derived from the research question, but the awareness of the rich toolkit available can sometimes inspire researchers to ask questions that are bold and transformative (S. C. Schmidt & Marwick, 2020). Here I will leverage the extensive corpus in experimental designs and inferences revolving around stone artefacts to clarify its meaning and demonstrate the necessity and potential of this framework.

2 What good is less-controlled experimentation?

The trade-off between causal inference (aka 'internal validity') and generalization (aka 'external validity') forms a central issue in experimental design across different disciplines (Eren et al., 2016; Roe & Just, 2009: 1266-1267). Even in fields known for their development of rigorous and well-controlled experimental methods such as cognitive psychology and neuroscience, researchers have started to use relatively naturalistic stimuli more frequently and advocate a paradigm shift to semi-controlled experiments due to the generalizability crisis, namely the prevailing mismatch between phenomenon of interest and measured variables in psychological science (Nastase et al., 2020; Shamay-Tsoory & Mendelsohn, 2019; Sonkusare et al., 2019; Yarkoni, 2022). In contrast, the past decades have witnessed experimental archaeology's growing research interests focusing on the causal mechanism at the behavioural level in the explanation of material culture variation (Eren et al., 2016; Lin et al., 2018). In the context of stone artefact replication, one typical research design emphasizing causality over generalizability is the use of knapping machines/robots (Li et al., 2022; Pfleging et al., 2019), which has helped map out the physical constraints of stone artefact manufacture and use through the identification of causal relationships between input (force, exterior platform angle, platform depth, etc.) and outcome variables (flake size, flake shape, wear formation, etc.). All variables of interest in this setting are relatively easy to measure, quantify, and control, but this type of design can be insufficient in inferring how context-generic principles interact in a particular context as reflected in real-world conditions. This research orientation essentially prioritize the material science aspect over the social science aspect of experimental archaeology. Similarly, standardized artificial materials like bricks (Lombao et al., 2017) or foam blocks (Schillinger et al., 2016) have been used to standardize materials and/or reduce learning demands in experimental studies focusing on the transmission of lithic technologies, with implications for the generalizability of results (Liu et al., 2023). In real-world knapping, each rock has a different shape and often different physical properties such as inner cracks and inclusions, and this heterogeneity itself represents a critical variable in cultural transmission and skill development (Proffitt et al., 2022).

On the other hand, less-controlled experiments, which have been traditionally known as naturalistic or actualistic expeirments (Conrad et al., 2023; see terminological critiques of Eren & Meltzer, 2024), pay more attention to how experimental insights can be generalized to archaeological samples by incorporating authentic materials and plausible social settings with a certain degree of compromised control (Outram, 2008). Back to the cases of cultural transmission experiments, a less-controlled experiment would involve the use of natural rocks with varied morphology instead of standardized artificial materials as well as human demonstrators instead of videos of knapping instruction, despite the fact that the latter will remain consistent across individuals. Unlike strictly controlled experiments testing one variable of interest each time, variation could be easily observed in less-controlled experiments by design. This feature is crucial and cannot be simply replaced by ethnographic records, because many palaeolithic technological components do not have analogues in contemporary non-industrial societies (Arthur, 2018; e.g., Stout, 2002). While uncontrolled variation has traditionally been viewed as highly problematic, statistical techniques for developing causal inference from observational data, of the kind produced by less-controlled experiments, have also been greatly boosted in epidemiology and economics in recent years (Cunningham, 2021; Hernan & Robins, 2023). Naturalistic experiment can serve a heuristic role in hypothesis generation, aligning with the perspective of Lin et al. (2018: 680-681), who proposed that the interaction between less-controlled and strictly controlled experiment "operates in a cyclical form of induction and deduction."

3 Many places, many voices

Contemporary practices in experimental archaeology, as manifested by the fact that a majority of scholarly publications are produced as results of experiments conducted by a single knapper with a dual identity of researcher (Whittaker, 2004), tend to be restrained by the cognitive bias known as the 'curse of knowledge' or 'curse of expertise'. This psychological term originally refers to the phenomenon that it is extremely challenging for experts to ignore the information that is held by them but not others, particularly novices (Hinds, 1999), but it has further implications for the sample representativeness in experimental archaeology. When the knapping expertise is gradually formed through multiple years of observations and trial-and-error learning, an expert knapper develops some specific ways of strategic planning, motor habits (and their associated impacts on anatomical forms like wrist and elbow), preferences of percussor and raw material types, as well as familiarity of various techniques that become unforgettable (Moore, 2020: 654). The existence of this cognitive bias is not inherently bad, and these many years of experience should be appreciated and celebrated by experimental archaeologists. However, what is problematic is that the results of replication experiments conducted by these experienced practitioners, often in settings of single knapper, has been constantly framed as generalization regarding the evolution of technology and cognition that masks a vast range of technological diversity.

Modern flintknapping techniques, as a research subject and a scientific method, originated from hobbyists' individualistic trials of reverse engineering during the 19th century (Coles, 1979; Flenniken, 1984; Johnson, 1978; Whittaker, 1994: 54-61). Hobbyist knappers represent a huge repertoire of technological knowledge that does not fully overlap with what is acquired by academic knappers. They tend to generate ideas that may appear to be counter-intuitive at first glance for academics. One such example is the utility of obtuse edge angle as demonstrated by Don Crabtree (1977), a mostly self-educated flintknapper yet one of the most important figures in experimental archaeology. In his experiment, Crabtree demonstrated the excellent performance of blade dorsal ridge on tasks like shaving and cutting hard materials, challenging the traditional perspective on producing sharp lateral edges as the sole purpose of stone toolmaking and shedding light on future functional reconstruction through the use-wear analysis. It is rather unfortunate that collaborations between academics and hobbyists are less common than expected due to their complicated and uneasy relationships as detailed in Whittaker's (2004) ethnography. Likewise, novices' lack of expertise also helps to mitigate the 'curse of knowledge' bias that may hinder expert knappers. Their involvement can potentially lead to the discovery of alternative methods, techniques, and interpretations that may have been overlooked by experts.

Emphasizing variation at its core, the Triple P conceptual framework recognizes that experimental archaeology can greatly benefit from diverse perspectives (Pargeter et al., 2023: 164) and thereby inherently adopts a collaborative mode of knowledge production, which has been recently advocated in experimental studies (Liu & Stout, 2023; Ranhorn et al., 2020) and museum collection studies (Timbrell, 2023) of stone artefacts. Furthermore, the Triple P framework acknowledges that communities living in specific geographical areas possess unique insights and understanding of their cultural heritage. This emphasis on team efforts and inclusivity allows for a more complete understanding of the non-utilitarian or unexpected aspects of raw material procurement (Batalla, 2016) and selection (Arthur, 2021), pre-treatment (Maloney & Street, 2020), production (Griffin et al., 2013), and use (Martellotta et al., 2022; Milks et al., 2023) across different regions. Through ethical collaborations with those knapping practitioners in non-industrial societies in the research process, the framework allows their voices to be heard and their contributions to be acknowledged. This not only enhances the quality of research outcomes but also fosters a sense of ownership and pride within these communities, strengthening the connection between archaeological research and the people it directly affects (Douglass, 2020; Marshall, 2002; Montgomery & Fryer, 2023).

However, the facilitation of large-scale collaborations faces challenges within the current system of research evaluation. The prevailing practice of attributing credit primarily to the first author and senior (last/corresponding) author in peer-reviewed journal papers hampers the recognition of multiple contributors. This system often overlooks the valuable input of collaborators who may not fit into the traditional authorship structure but have made significant intellectual and practical contributions to the research. To truly embrace the principles of collaboration and inclusivity, there is a need for a reevaluation of the research evaluation system, allowing for proper acknowledgment of the diverse voices and contributions involved in large-scale collaborations.

4 The Triple P framework in action

The implementation of the Triple P framework involves the collection of process-level (ethological) and perception-level (ethnographic) data (**Figure** 1), which is critical to address equifinality and multifinality (Hiscock, 2004; Nami, 2010; Premo, 2010), two daunting challenges in archaeological inference. Equifinality refers to situations in which a similar state or consequence can be achieved through multiple different paths, while multifinality emerges when a similar process can lead to multiple ends. While we cannot fully solve these two problems and accurately reconstruct the past behavioural processes simply based on materials remains, context-rich experiments involving the collection of ethological and ethnographic data can help us better document an enlarged range of possible combinations of variation and draw a more informed inference (Reynolds, 1999). The importance of specifying and documenting the context information of both the experiment and the phenomenon of interest has also been recently highlighted in psychological sciences (Holleman et al., 2020).

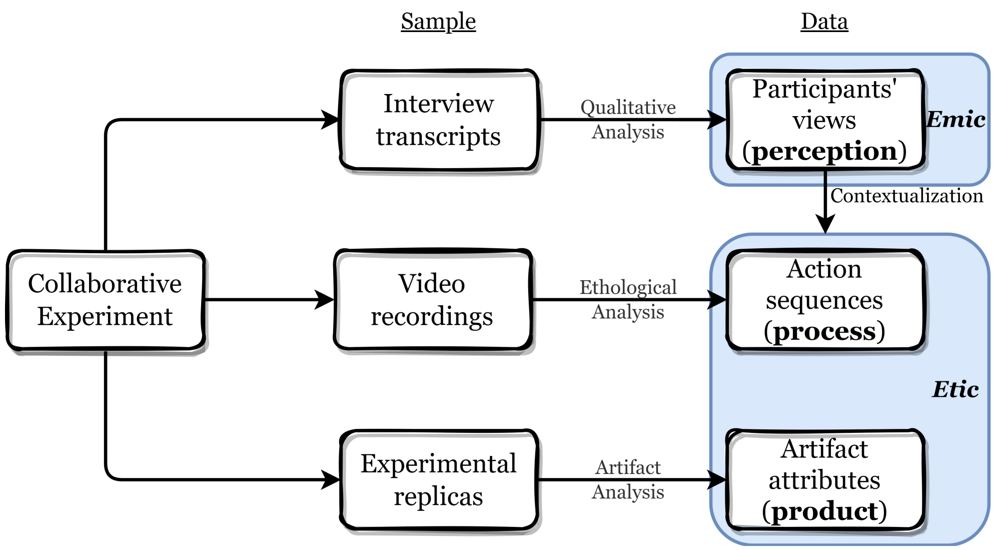

Figure 1: A schematic diagram demonstrating how to operationalize the Perception-Process-Product conceptual framework.

4.1 Product-level data

Traditionally speaking, the product-level data, namely the documentation and analysis of replicas, form the sole research subject of experimental archaeology and serve as the tangible foundation for analogical inference in the interpretation of archaeological materials. It can exist in the form of spreadsheets containing detailed technological attributes, photos and illustrations, or high-resolution 3D scans of individual artefacts or a whole assemblage. No particular modification regarding the collection procedure of product-level data is required in the context of the Triple P framework, although the definition of variables measured and the documentation techniques (models of camera/scanners, light setting, processing software version and workflow, etc.) should be always available in the relevant meta-data. I also strongly recommend adopting good habits in spreadsheet data organization (Broman & Woo, 2018).

4.2 Process-level data

While systematic behavioural coding methods widely used in the study of non-human animal behaviour (Fragaszy & Mangalam, 2018) are still largely neglected among archaeologists, attempts to reconstruct behavioural sequences involved in the manufacture of material remains are not infrequent. One such example is the cognigram, which was first systematically developed and applied in archaeological research by Haidle (Haidle, 2009, 2010, 2014, 2023). A cognigram is a graphical representation of the reconstructed behaviour behind archaeological artefacts in chronological order of appearance (Haidle, 2014), which essentially represents an abstracting process of a series of action sequences achieving a similar goal. This approach provides an elegant descriptive methodology yet is limited by its normative and analytical orientation, meaning it cannot handle variation very well. To some extent, it describes the minimal steps to achieve a goal from the perspective of reverse engineering and reflects the analyst's own causal perception. However, this may be biased because 1) certain causal insights in stone fracture mechanics remain opaque to academic knappers until they are revealed through controlled experiments by Dibble and his colleagues (Li et al., 2022) 2) ethnographic studies demonstrated that expert non-academic practitioners can have a different set of causal understanding (Harris et al., 2021).

Consequently, we need to accumulate more real-world data by recording a large number of toolmaking videos and conducting systematic ethogram analysis. With the emergence of new software platforms such as BORIS (Friard & Gamba, 2016), the difficulty of coding has decreased significantly in recent years (**Figure** 2). Here I use a modified version of action grammar developed by (Stout et al., 2021) as an example, among multiple coding schemes featuring different research focus (Muller et al., 2023) or granularity (Cueva-Temprana et al., 2019; Mahaney, 2014; Roux & David, 2005). The knapping action recorded in videos can be coded following the ethogram presented in **Table** 1. Depending on the original research question, sequences of coded actions can then be used in further analysis, such as complexity (Stout et al., 2021), similarity (Cristino et al., 2010; Mobbs et al., 2021), etc.

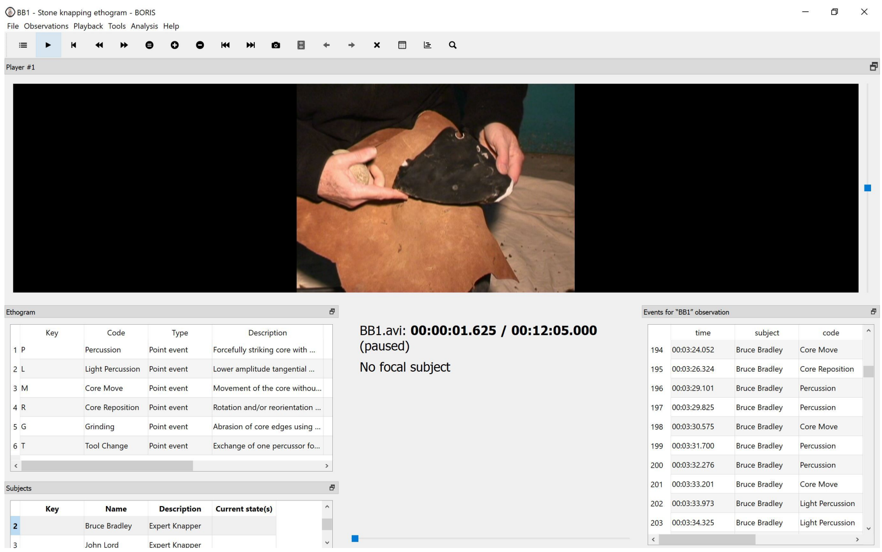

Figure 2: An example of coding a handaxe knapping session using the BORIS software.

Table 1: A modified version of the original action grammar presented in (Stout et al., 2021)

Action	Definition
Percussion	Forcefully striking core with percussor (hammerstone or antler billet) in such a way as to potentially remove a flake
Light Percussion	Lower amplitude tangential strike to the tool edge of the kind often employed for platform preparation
Core Move	Movement of the core without a change in grip, which often occurs during the core inspection
Core Reposition	Rotation and/or reorientation of the core involving repositioning of the hand, which is often associated with the transition to a new percussion target
Grinding	Abrasion of core edges using a hammerstone. The abrasion movement should come from at least two different directions.
Tool Change	Exchange of one percussor for another
Winding Up	Preparational percussor movements towards the core that do not lead to the detachment of flakes, which can either be in direct contact with cores or not.

4.3 Perception-level data

Ethnographies revolving around experimental archaeology as a field (Reeves Flores, 2012), as well as practices of specific technologies like flintknapping, including contemporary U.S. hobbyists (Whittaker, 2004) and knapping practitioners in various non-industrial societies (Arthur, 2018; Stout, 2002), are far from novel. However, ethnography has never been formally recognized as a legitimate research method in experimental archaeology. Echoing with the recent trends of adopting embodied cognition (Varela et al., 2017) in archaeological research (Malafouris, 2013), ethnographic data and methods can reveal hidden information (e.g., intention, phenomenology) that is otherwise irretrievable and thus should occupy a unique niche in experimental archaeology. Within the broader context of burgeoning interest in mixed-method research in contemporary social science (Creswell & Clark, 2017), this also echoes the post-positivist turn in psychology in the past decades, particularly the emphasis on the value of incorporating qualitative research (Stout, 2021; Syed & McLean, 2022; Weger et al., 2019).

Through participant observation, interviews, and detailed field notes, ethnography can capture the subtle nuances of perception, such as cognitive affordances (Hussain & Will, 2021; Roepstorff, 2008), sensory experiences, social interactions, and cultural meanings associated with the experimental activities (Gowlland, 2019). Compared with the ethological methods, the interview questions and participant observation in ethnographic methods feature an even higher degree of freedom and rely more heavily on the research question as well as ad-hoc interaction. One potential application of ethnographic methods in experimental archaeology of stone artefacts is asking knappers about the intentions of each action and see how it matches with the results as revealed by lithic analysis of replicas, which can provide crucial contextual information addressing the issues of equifinality and multifinality in the formation of lithic assemblage. Instead of seeing intention as something abstruse or unapproachable in archaeology (David, 2004; Russell, 2004), the Triple P framework adopts a novel definition proposed by Quillien and German (2021: 1) from the perspective of causal perception, namely 'an agent did X *intentionally* to the extent that X was causally dependent on how much the agent wanted X to happen (or not to happen).' In this sense, the mismatch between how different individuals perceive cause-and-effect relationships and how they are organized according to physical laws is exactly where interesting variation emerges and where ethnography become necessary.

4.4 Multi-level data curation

The comparative study and large-scale synthesis of variation data require the building of centralized, open-access, and carefully curated data infrastructure (Marwick & Birch, 2018), which unfortunately still does not exist yet in experimental archaeology. The accessibility and availability of experimental data can foster collaboration and enhance the reproducibility and transparency of research findings, as others can verify and validate the results by examining the original data. Moreover, a centralized database also promotes data preservation and long-term accessibility. By storing experimental data in a structured and organized manner, it safeguards valuable information from potential loss or degradation over time. This preservation ensures that the data remains accessible for future researchers, avoiding the loss of valuable insights and preventing the need for redundant and costly repetitions of experiments. It also allows for the reanalysis of existing data, facilitating discoveries and insights that may not have been initially anticipated. However, it has been widely acknowledged that the reuse of archaeological data has not received enough attention among researchers in our discipline (Faniel et al., 2018; Huggett, 2018; Moody et al., 2021).

Among the three dimensions of the Triple P framework, the product-level data are usually stored in the format of spreadsheets, photos, and 3D models, and the perception-level data formats mainly include audio files and their transcribed texts, whereas videos are the main vector of process-level data, a rather non-traditional data format in archaeological research featuring the largest file size compared with the other two. As such, following data sharing principles of FAIR (Wilkinson et al., 2016) and CARE (Carroll et al., 2020), the Triple P framework recommends Databrary (Gilmore et al., 2015; Simon et al., 2015), a web-based library originally designed for developmental scientists, as the main data curation platform, where researchers can freely upload video files with no size limit and related metadata that can connect with different types of data within the same project. Databrary has three advantages compared with other data storage solutions: a) no cost from the side of researcher; 2) long-term data security monitored by a specialized maintenance team; and 3) fostering potential collaborations between experimental archaeologists and developmental psychologists.

5 Conclusion

Through the broadening of traditional data types and recording methods revolving around experimental replicas *per se*, the Triple P conceptual framework allows the amplified multiscale expression of material cultural variation. It is also compatible with many theoretical orientations, ranging from behavioural archaeology (emphasis on video recording of behavioural processes) through evolutionary archaeology (emphasis on the amplification of variation) to post-processual archaeology (emphasis on perception through ethnography). In terms of its research practice, it embraces a collaborative mode of knowledge production by involving a more diverse pool of stakeholders. The innovativeness, flexibility, and inclusiveness of the Triple P conceptual framework has enormous potential in redefining what can be and what should be studied by experimental archaeology as a field and thereby contributing to a better understanding of our deep past.

6 Acknowledgements

I thank Dietrich Stout for his helpful comments on earlier drafts of this article and Mark Moore for an inspiring discussion. This study was supported by the Leakey Foundation research grant titled 'Inferring skill reproduction from stone artifacts: A middle-range approach' as well as the International Society of Human Ethology's Owen Aldis Award project titled 'Sealed in stones: The computational ethology of stone toolmaking and its implications to the evolution of cultural transmission.'

References

Acerbi, A., Enquist, M., & Ghirlanda, S. (2009). Cultural evolution and individual development of openness and conservatism. *Proceedings of the National Academy of Sciences*, *106*(45), 18931–18935. https://doi.org/10.1073/pnas.0908889106

Antón, S. C., & Kuzawa, C. W. (2017). Early homo, plasticity and the extended evolutionary synthesis. *Interface Focus*, *7*(5), 20170004. https://doi.org/10.1098/rsfs.2017.0004

Apicella, C., Norenzayan, A., & Henrich, J. (2020). Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior*, *41*(5), 319–329. https://doi.org/10.1016/j.evolhumbehav.2020.07.015

Arthur, K. W. (2018). *The lives of stone tools: Crafting the status, skill, and identity of flintknappers*. University of Arizona Press.

Arthur, K. W. (2021). Material Scientists: Learning the Importance of Colour and Brightness from Lithic Practitioners. *Cambridge Archaeological Journal*, *31*(2), 293–304. https://doi.org/10.1017/S0959774320000347

Barrett, H. C. (2020). Towards a Cognitive Science of the Human: Cross-Cultural Approaches and Their Urgency. *Trends in Cognitive Sciences*, *24*(8), 620–638. https://doi.org/10.1016/j.tics.2020.05.007

Batalla, A. N. (2016). Studies of indigenous lithic procurement in Uruguay and their implications for Southern Cone archaeology. *Journal of Lithic Studies*, *3*(1), 265–292. https://doi.org/10.2218/jls.v3i1.1522

Bettinger, R. L., & Baumhoff, M. A. (1982). The Numic Spread: Great Basin Cultures in Competition. *American Antiquity*, *47*(3), 485–503. https://doi.org/10.2307/280231

Blessing, M. A., & Schmidt, P. (2021). On the efficiency of palaeolithic birch tar making. *Journal of Archaeological Science: Reports*, *38*, 103096. https://doi.org/10.1016/j.jasrep.2021.103096

Broman, K. W., & Woo, K. H. (2018). Data organization in spreadsheets. *The American Statistician*, *72*(1), 2–10. https://doi.org/10.1080/00031305.2017.1375989

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, *19*(1), 43. https://doi.org/10.5334/dsj-2020-043

Clay, Z., & Tennie, C. (2018). Is Overimitation a Uniquely Human Phenomenon? Insights From Human Children as Compared to Bonobos. *Child Development*, *89*(5), 1535–1544. https://doi.org/10.1111/cdev.12857

Coles, J. M. (1979). *Experimental archaeology*. Academic Press.

Conrad, G., Hough, S., Baldino, J., Gala, N., Buchanan, B., Walker, R. S., Key, A., Redmond, B. G., Bebber, M. R., & Eren, M. I. (2023). Clovis bone versus stone weapon tip penetration: Thinking about relative costs and benefits, experimental assumptions, and archaeological unknowns at sheriden cave, ohio, u.s.a. *Journal of Archaeological Science: Reports*, *52*, 104295. https://doi.org/10.1016/j.jasrep.2023.104295

Crabtree, D. E. (1977). *The obtuse angle as a functional edge* (D. Ingersoll, J. E. Yellen, & W. MacDonald, Eds.; pp. 38–51). Columbia University Press.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research* (3rd edition). SAGE Publications, Inc.

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692–700. https://doi.org/10.3758/BRM.42.3.692

Cueva-Temprana, A., Lombao, D., Morales, J. I., Geribàs, N., & Mosquera, M. (2019). Gestures during knapping: A two-perspective approach to pleistocene technologies. *Lithic Technology*, *44*(2), 74–89. https://doi.org/10.1080/01977261.2019.1587255

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press. https://doi.org/10.2307/j.ctv1c29t27

David, B. (2004). Intentionality, Agency and an Archaeology of Choice. *Cambridge Archaeological Journal*, *14*(1), 67–71. https://doi.org/10.1017/S0959774304220054

Domínguez-Rodrigo, M. (2008). Conceptual premises in experimental design and their bearing on the use of analogy: An example from experiments on cut marks. *World Archaeology*, *40*(1), 67–82. https://doi.org/10.1080/00438240701843629

Douglass, K. (2020). Amy ty lilin-draza'ay: Building Archaeological Practice on Principles of Community. *African Archaeological Review*, *37*(3), 481–485. https://doi.org/10.1007/s10437-020-09404-8

Eren, M. I., Lycett, S. J., Patten, R. J., Buchanan, B., Pargeter, J., & O'Brien, M. J. (2016). Test, model, and method validation: The role of experimental stone artifact replication in hypothesis-driven archaeology. *Ethnoarchaeology: Journal of Archaeological, Ethnographic and Experimental Studies*, *8*(2), 103–136. https://doi.org/10.1080/19442890.2016.1213972

Eren, M. I., & Meltzer, D. J. (2024). Controls, conceits, and aiming for robust inferences in experimental archaeology. *Journal of Archaeological Science: Reports*, *53*, 104411. https://doi.org/10.1016/j.jasrep.2024.104411

Faniel, I. M., Austin, A., Kansa, E., Kansa, S. W., France, P., Jacobs, J., Boytner, R., & Yakel, E. (2018). Beyond the Archive: Bridging Data Creation and Reuse in Archaeology. *Advances in Archaeological Practice*, *6*(2), 105–116. https://doi.org/10.1017/aap.2018.2

Flenniken, J. J. (1984). The past, present, and future of flintknapping: An anthropological perspective. *Annual Review of Anthropology*, *13*(1), 187–203. https://doi.org/10.1146/annurev.an.13.100184.001155

Fragaszy, D. M., & Mangalam, M. (2018). *Chapter Five - Tooling* (M. Naguib, L. Barrett, S. D.

Healy, J. Podos, L. W. Simmons, & M. Zuk, Eds.; Vol. 50, pp. 177–241). Academic Press. https://doi.org/10.1016/bs.asb.2018.01.001

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330. https://doi.org/10.1111/2041-210X.12584

Gergely, G., & Csibra, G. (2006). *Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge* (S. C. Levinson & N. J. Enfield, Eds.; pp. 229–255). Berg Publishers.

Ghirlanda, S., Enquist, M., & Nakamaru, M. (2006). Cultural evolution develops its own rules: The rise of conservatism and persuasion. *Current Anthropology*, *47*(6), 1027–1034. https://doi.org/10.1086/508696

Gilmore, R., Adolph, K., Millman, D., Steiger, L., & Simon, D. (2015). Sharing displays and data from vision science research with databrary. *Journal of Vision*, *15*(12), 280. https://doi.org/10.1167/15.12.280

Gowlland, G. (2019). The sociality of enskilment. *Ethnos*, *84*(3), 508–524. https://doi.org/10.1080/00141844.2018.1455726

Griffin, D., Freedman, D. L., Nicholson, B., McConachie, F., & Parmington, A. (2013). The koorong project: Experimental archaeology and wurundjeri continuation of cultural practices. *Excavations, Surveys and Heritage Management in Victoria*, *2*, 5965.

Haidle, M. N. (2009). How to think a simple spear. In S. A. de Beaune, F. L. Coolidge, & T. Wynn (Eds.), *Cognitive archaeology and human evolution* (pp. 57–73). Cambridge University Press.

Haidle, M. N. (2010). Working-memory capacity and the evolution of modern cognitive potential: Implications from animal and early human tool use. *Current Anthropology*, *51*(S1), S149–S166. https://doi.org/10.1086/650295

Haidle, M. N. (2014). Building a bridge—an archeologist's perspective on the evolution of causal cognition. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01472

Haidle, M. N. (2023). *Cognigrams: Systematically reconstructing behavioral architectures as a basis for cognitive archaeology* (T. Wynn, K. A. Overmann, & F. L. Coolidge, Eds.; p. C12S1C12S8). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780192895950.013.12

Harris, J. A., Boyd, R., & Wood, B. M. (2021). The role of causal knowledge in the evolution of traditional technology. *Current Biology*, *31*(8), 1798–1803.e3. https://doi.org/10.1016/j.cub.2021.01.096

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review*, *91*(2), 73–78. https://doi.org/10.1257/aer.91.2.73

Hernan, M. A., & Robins, J. M. (2023). *Causal inference: What if.* CRC Press.

Heyes, C. (2012). Simple minds: A qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2695–2703. https://doi.org/10.1098/rstb.2012.0217

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, *5*, 205–221. https://doi.org/10.1037/1076-898X.5.2.205

Hiscock, P. (2004). Slippery and Billy: Intention, Selection and Equifinality in Lithic Artefacts. *Cambridge Archaeological Journal*, *14*(1), 71–77. https://doi.org/10.1017/S0959774304230050

Holleman, G. A., Hooge, I. T., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach'and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, *11*, 721.

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. https://doi.org/10.1007/s10071-004-0239-6

Huggett, J. (2018). Reuse Remix Recycle: Repurposing Archaeological Digital Data. *Advances in Archaeological Practice*, *6*(2), 93–104. https://doi.org/10.1017/aap.2018.1

Hussain, S. T., & Will, M. (2021). Materiality, Agency and Evolution of Lithic Technology: an Integrated Perspective for Palaeolithic Archaeology. *Journal of Archaeological Method and Theory*, *28*(2), 617–670. https://doi.org/10.1007/s10816-020-09483-6

Johnson, L. L. (1978). A history of flint-knapping experimentation, 1838-1976 [and comments and reply]. *Current Anthropology*, *19*(2), 337–372. https://doi.org/10.1086/202078

Kafaee, M., Daviran, E., & Taqavi, M. (2022). The QWERTY keyboard from the perspective of the Collingridge dilemma: lessons for co-construction of human-technology. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01573-1

Li, L., Lin, S. C., McPherron, S. P., Abdolahzadeh, A., Chan, A., Dogandžić, T., Iovita, R., Leader, G. M., Magnani, M., Rezek, Z., & Dibble, H. L. (2022). A Synthesis of the Dibble et al. Controlled Experiments into the Mechanics of Lithic Production. *Journal of Archaeological Method and Theory*. https://doi.org/10.1007/s10816-022-09586-2

Lin, S. C., Rezek, Z., & Dibble, H. L. (2018). Experimental Design and Experimental Inference in Stone Artifact Archaeology. *Journal of Archaeological Method and Theory*, *25*(3), 663–688. https://doi.org/10.1007/s10816-017-9351-1

Liu, C., Khreisheh, N., Stout, D., & Pargeter, J. (2023). Differential effects of knapping skill acquisition on the cultural reproduction of Late Acheulean handaxe morphology: Archaeological and experimental insights. *Journal of Archaeological Science: Reports*, *49*, 103974. https://doi.org/10.1016/j.jasrep.2023.103974

Liu, C., & Stout, D. (2023). Inferring cultural reproduction from lithic data: A critical review. *Evolutionary Anthropology: Issues, News, and Reviews*, *32*(2), 83–99. https://doi.org/10.1002/evan.21964

Lombao, D., Guardiola, M., & Mosquera, M. (2017). Teaching to make stone tools: new experimental evidence supporting a technological hypothesis for the origins of language. *Scientific Reports*, *7*(1), 14394. https://doi.org/10.1038/s41598-017-14322-y

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. https://doi.org/10.1073/pnas.0704452104

Mahaney, R. A. (2014). Exploring the complexity and structure of acheulean stoneknapping in relation to natural language. *PaleoAnthropology*, *2014*, 586606. https://doi.org/10.4207/PA.2014.ART90

Malafouris, L. (2013). *How things shape the mind: A theory of material engagement*. The MIT Press.

Maloney, T. R., & Street, M. (2020). Hot debate: Identifying heat treatment in Australian archaeology using science and modern indigenous knowledge. *Quaternary Science Reviews*, *241*, 106431. https://doi.org/10.1016/j.quascirev.2020.106431

Marshall, Y. (2002). What is community archaeology? *World Archaeology*, *34*(2), 211–219. https://doi.org/10.1080/0043824022000007062

Martellotta, E. F., Perston, Y. L., Craft, P., Wilkins, J., & Langley, M. C. (2022). Beyond the main function: An experimental study of the use of hardwood boomerangs in retouching activities. *PLOS ONE*, *17*(8), e0273118. https://doi.org/10.1371/journal.pone.0273118

Marwick, B., & Birch, S. E. P. (2018). A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing. *Advances in Archaeological Practice*, *6*(2), 125–143. https://doi.org/10.1017/aap.2018.3

Mesoudi, A., & O'Brien, M. J. (2008). The cultural transmission of great basin projectile-point technology i: An experimental simulation. *American Antiquity*, *73*(1), 328. https://doi.org/https://doi.org/10.1017/S0002731600041263

Milks, A., Hoggard, C., & Pope, M. (2023). Reassessing the Interpretative Potential of Ethnographic Collections for Early Hunting Technologies. *Journal of Archaeological Method and Theory*. https://doi.org/10.1007/s10816-023-09635-4

Mindermann, S., & Armstrong, S. (2018). Occam's razor is insufficient to infer the preferences of irrational agents. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 5603–5614.

Mobbs, D., Wise, T., Suthana, N., Guzmán, N., Kriegeskorte, N., & Leibo, J. Z. (2021). Promises and challenges of human computational ethology. *Neuron*, *109*(14), 2224–2238. https://doi.org/10.1016/j.neuron.2021.05.021

Montgomery, L. M., & Fryer, T. C. (2023). The future of archaeology is (still) community collaboration. *Antiquity*, *97*(394), 795–809. https://doi.org/10.15184/aqy.2023.98

Moody, B., Dye, T., May, K., Wright, H., & Buck, C. (2021). Digital chronological data reuse in archaeology: Three case studies with varying purposes and perspectives. *Journal of Archaeological Science: Reports*, *40*, 103188. https://doi.org/10.1016/j.jasrep.2021.103188

Moore, M. W. (2020). Hominin Stone Flaking and the Emergence of 'Top-down' Design in Human Evolution. *Cambridge Archaeological Journal*, *30*(4), 647–664. https://doi.org/10.1017/S0959774320000190

Morin, O. (2022). Cultural Conservatism. *Journal of Cognition and Culture*, *22*(5), 406–420. https://doi.org/10.1163/15685373-12340142

Muller, A., Shipton, C., & Clarkson, C. (2023). The Proceduralization of Hominin Knapping Skill: Memorizing Different Lithic Technologies. *Cambridge Archaeological Journal*, 1–18. https://doi.org/10.1017/S0959774323000070

Nami, H., G. (2010). Theoretical Reflections on Experimental Archaeology and Lithic Technology: Issues on Actualistic Stone Tools Analysis and Interpretation. In H. Nami G. (Ed.), *Experiments and Interpretation of Traditional Technologies: Essays in Honor of Errett Callahan* (pp. 91–168). Ediciones de ArqueologÌa Contempornea.

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman Children and the Origins of Human Cultural Cognition. *Psychological Science*, *21*(5), 729–736. https://doi.org/10.1177/0956797610368808

Outram, A. K. (2008). Introduction to experimental archaeology. *World Archaeology*, *40*(1), 1–6. https://www.jstor.org/stable/40025310

Pargeter, J., Liu, C., Kilgore, M. B., Majoe, A., & Stout, D. (2023). Testing the Effect of Learning Conditions and Individual Motor/Cognitive Differences on Knapping Skill Acquisition. *Journal of Archaeological Method and Theory*, *30*(1), 127–171. https://doi.org/10.1007/s10816-022-09592-4

Pfleging, J., Iovita, R., & Buchli, J. (2019). Influence of force and duration on stone tool wear: results from experiments with a force-controlled robot. *Archaeological and Anthropological Sciences*, *11*(11), 5921–5935. https://doi.org/10.1007/s12520-018-0729-0

Premo, L. S. (2010). Equifinality and explanation: Thoughts on the role of agent-based modeling in postpositivist archaeology. In A. Costopoulos & M. W. Lake (Eds.), *Simulating Change: Archaeology Into the Twenty-first Century* (pp. 28–37). University of Utah Press.

Proffitt, T., Bargalló, A., & Torre, I. de la. (2022). The Effect of Raw Material on the Identification of Knapping Skill: a Case Study from Olduvai Gorge, Tanzania. *Journal of Archaeological Method and Theory*, *29*(1), 50–82. https://doi.org/10.1007/s10816-021-09511-z

Quillien, T., & German, T. C. (2021). A simple definition of 'intentionally'. *Cognition*, *214*, 104806. https://doi.org/10.1016/j.cognition.2021.104806

Ranhorn, K. L., Pargeter, J., & Premo, L. S. (2020). Investigating the evolution of human social learning through collaborative experimental archaeology. *Evolutionary Anthropology: Issues, News, and Reviews*, *29*(2), 53–55. https://doi.org/10.1002/evan.21823

Reeves, D., Bury, R., & Robinson, D. W. (2009). Invoking occam's razor: Experimental pigment processing and an hypothesis concerning emigdiano chumash rock art. *Journal of California and Great Basin Anthropology*, *29*(1), 59–67. https://www.jstor.org/stable/27825902

Reeves Flores, J. (2012). *Experimental archaeology: an ethnography of its perceived value and impact in archaeological research* [PhD thesis]. https://ore.exeter.ac.uk/repository/handle/10871/9041

Reynolds, P. J. (1999). The nature of experiment in archaeology. In A. Harding (Ed.), *Experiment and design: Archaeological studies in honour of John Coles* (pp. 156–162). Oxbow Books.

Roe, B. E., & Just, D. R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American*

Journal of Agricultural Economics, *91*(5), 1266–1271. https://www.jstor.org/stable/20616293

Roepstorff, A. (2008). Things to think with: Words and objects as material symbols. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2049–2054. https://doi.org/10.1098/rstb.2008.0015

Roux, V., & David, É. (2005). *Planning abilities as a dynamic perceptual-motor skill: an actualist study of different levels of expertise involved in stone knapping* (V. Roux & B. Bril, Eds.; pp. 91–108). McDonald Institute for Archaeological Research. https://shs.hal.science/halshs-00120262

Russell, L. (2004). Drinking from the Penholder: Intentionality and Archaeological Theory. *Cambridge Archaeological Journal*, *14*(1), 64–67. https://doi.org/10.1017/S0959774304210058

Schiffer, M. B. (2010). *Behavioral Archaeology: Principles and Practice*. Routledge.

Schillinger, K., Mesoudi, A., & Lycett, S. J. (2016). Copying error, evolution, and phylogenetic signal in artifactual traditions: An experimental approach using "model artifacts". *Journal of Archaeological Science*, *70*, 23–34. https://doi.org/10.1016/j.jas.2016.04.013

Schmidt, P., Blessing, M., Rageot, M., Iovita, R., Pfleging, J., Nickel, K. G., Righetti, L., & Tennie, C. (2019). Birch tar production does not prove Neanderthal behavioral complexity. *Proceedings of the National Academy of Sciences*, *116*(36), 17707–17711. https://doi.org/10.1073/pnas.1911137116

Schmidt, S. C., & Marwick, B. (2020). Tool-Driven Revolutions in Archaeological Science. *Journal of Computer Applications in Archaeology*, *3*(1), 1832. https://doi.org/10.5334/jcaa.29

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on Psychological Science*, *14*(5), 841–859. https://doi.org/10.1177/1745691619856350

Simon, D. A., Gordon, A. S., Steiger, L., & Gilmore, R. O. (2015). *Databrary: Enabling sharing and reuse of research video*. 279280. https://doi.org/10.1145/2756406.2756951

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*, *23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Stengelin, R., Hepach, R., & Haun, D. B. M. (2020). Cross-cultural variation in how much, but not whether, children overimitate. *Journal of Experimental Child Psychology*, *193*, 104796. https://doi.org/10.1016/j.jecp.2019.104796

Stout, D. (2002). Skill and cognition in stone tool production: An ethnographic case study from

irian jaya. *Current Anthropology*, *43*(5), 693–722. https://doi.org/10.1086/342638
Stout, D. (2021). The cognitive science of technology. *Trends in Cognitive Sciences*, *25*(11), 964–977. https://doi.org/10.1016/j.tics.2021.07.005
Stout, D., Chaminade, T., Apel, J., Shafti, A., & Faisal, A. A. (2021). The measurement, evolution, and neural representation of action grammars of human behavior. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-92992-5
Syed, M., & McLean, K. C. (2022). Disentangling paradigm and method can help bring qualitative research to post-positivist psychology and address the generalizability crisis. *Behavioral and Brain Sciences*, *45*, e32. https://doi.org/10.1017/S0140525X21000431
Timbrell, L. (2023). A Collaborative Model for Lithic Shape Digitization in Museum Settings. *Lithic Technology*, *48*(1), 31–42. https://doi.org/10.1080/01977261.2022.2092299
Varela, F. J., Thompson, E., & Rosch, E. (2017). *The Embodied Mind: Cognitive Science and Human Experience* (revised edition). The MIT Press.
Weger, U. W., Wagemann, J., & Tewes, C. (2019). Editorial: The challenges and opportunities of introspection in psychology: Theory and method. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02196
Whittaker, J. C. (1994). *Flintknapping: Making and Understanding Stone Tools*. University of Texas Press.
Whittaker, J. C. (2004). *American Flintknappers: Stone Age Art in the Age of Computers*. University of Texas Press.
Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18
Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685