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

Cheng Liu*

2023-09-07

Abstract

This paper presents the outline of the Perception-Process-Product ('Triple P') conceptual framework that aims to expand the scope of experimental archaeology. The Triple P framework emphasizes the amplification of multi-level variation and the identification of interacting relationships of variations across the levels of perception, process, and product. Here I propose the following three basic measures to put the Triple P framework into practice: 1) acknowledging the contribution and limitations of actualistic experiments properly; 2) encouraging collaborative projects that involves geographically diverse and non-traditional research participants such as hobbyists and novices; 3) adopting a workflow that normalize the collection and curation of ethological and ethnographic data in experimental projects.

Keywords: Experimental archaeology; Ethological analysis; Ethnographical analysis; The curse of knowledge; Collaborative knowledge production

Contents

*Department of Anthropology, Emory University, Atlanta, GA, USA; raylc1996@outlook.com

1 Introduction

This paper presents the Perception-Process-Product (hereinafter referred to as "Triple P") conceptual framework to expand the scope of experimental archaeology. Adopting the principle of Occam's razor (e.g., Domínguez-Rodrigo, 2008; Reeves et al., 2009), whether explicitly or implicitly, the field of experimental archaeology has long tended to center around the reverse engineering of a past technology in a minimal or least-effort manner while ignoring the rich contextual information it affords. Nevertheless, Occam's razor, this very cornerstone of our field, can be insufficient to infer the preferences of irrational agents possessing incomplete information (Armstrong & Mindermann, 2018) in tool design and use, and the two conditions described here, I believe, provide a better approximation of past humans displaying extensive cultural variation as opposed to the assumption of omniscient *Homo economicus* for most social scientists in the post-Kahneman era, but particularly anthropologists and archaeologists (Henrich et al., 2001).

As such, built upon early intellectual principles and practices in behavioral 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 behavioral channels (process) as well as sensory experiences (perception) and **b)** better identify the complex interacting relationships across these three levels of variations. To accomplish these two objectives, I advocate the following three measures as integral components of the Triple P framework: 1) acknowledging the contribution and limitations of actualistic experiments properly; 2) encouraging collaborative projects that involves geographically diverse and non-traditional research participants such as hobbyists and novices; 3) adopting a workflow that normalize the collection and curation of ethological and ethnographic data in experimental projects. It is no doubt that strategies of data collection and analysis of a given experimental project should be primarily derived from the research question, which can be legitimately narrow in scope, but the awareness of the rich toolkit available can sometimes inspire researchers to ask questions that are bold and transformative (Schmidt & Marwick, 2020). Here I will mainly leverage the extensive corpus in experimental designs and inferences revolving around stone artifacts to clarify its meaning and demonstrate the necessity and potential of this framework.

2 What good is actualistic experiment?

The trade-off between causality (aka "internal validity") and generalizability (aka "external validity") forms a central issue in experimental design (Eren et al., 2016; Roe & Just, 2009: 1266-1267), and the past decades have witnessed experimental archaeology's growing research interests focusing on the causal mechanism at the behavioral level in the explanation of material culture variation (Eren et al., 2016; Lin et al., 2018). In the context of stone artifact 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 artifact manufacture and use through the identification of multiple groups 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 is incapable of inferring how context-generic principles interact in a particular context as reflected in real-world conditions. In addition to the applications of machine knapping, the same problem is also incurred by the introduction of standardized artificial material like bricks (Lombao et al., 2017) or foam blocks (Schillinger et al., 2016) in experimental studies focusing on the transmission of lithic technologies, which was demonstrated to be problematic (Liu et al., 2023). In reality, 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 the process of skill acquisition. After all, these experimental results can only be as robust as their experimental settings.

More importantly, controlled experiments without randomization are not enough to infer causal mechanisms, which may be severely biased by factors such as individual differences (Pargeter et al., 2023), allocation concealment (Schulz & Grimes, 2002), and poor recruitment (Fletcher et al., 2012). This issue is rarely discussed in material science oriented experiments but should raise a critical concern in experimental archaeology as social science. Randomized Controlled Trials (RCT) *sensu stricto* as practiced in contemporary medicine and behavioral sciences, known for its high cost (e.g., Speich et al., 2019), are extremely rare in experimental archaeology when human participants were involved. Rather, most of our knowledge regarding the past is derived from data sets that can be characterized as Small, Unbalanced, Noisy, but Genuine (SUNG) (Arnaud et al., 2023) produced through experiments featuring small-sized convenience sample. It has also been a debatable issue whether Randomized Controlled Trial (RCT) represents the gold standard of knowledge in the realm of philosophy of science (Cartwright, 2007).

On the other hand, actualistic experiments 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, an actualistic 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. Interestingly, researchers in cognitive psychology and neuroscience, a field known for its development of rigorous and well-controlled experimental methods, have started to use naturalistic stimuli more frequently and advocate a paradigm shift to semi-controlled experiment (Nastase et al., 2020; Shamay-Tsoory & Mendelsohn, 2019; Sonkusare et al., 2019; Yarkoni, 2022). Unlike controlled experiments, variation could be easily observed in actualistic experiments by design. This feature is crucial and cannot be simply replaced by ethnographic records, because many paleolithic technological components are not displayed in contemporary non-industrial societies, which usually feature technological systems with groundstone artifacts as the target products (Arthur, 2018; e.g., Stout, 2002). Furthermore, statistical techniques for developing causal inference from observational data, which essentially represent the nature of results from actualistic experiments, have also been greatly boosted in epdedimiology and economics in recent years (Cunningham, 2021; Hernan & Robins, 2023; Nichols, 2007). Lastly, actualistic experiment can serve as a heuristic for hypothesis generation, aligning with the perspective of Lin et al. (2018: 680-681) and the argument put forth by Ingersoll and MacDonald (1977), who proposed that the interaction between actualistic and 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 (John C. Whittaker & Stafford, 1999), tend to be restrained by the cognitive bias known as the "curse of knowledge" or "curse of expertise". The curse of knowledge 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 (Camerer et al., 1989; Hinds, 1999). 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 experiences 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 huge 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, rather than from the inter-generational transmission of knapping knowledge spanning millions of years. This historical context is well elucidated by studies on the subject (Coles, 1979; Flenniken, 1984; Johnson, 1978; Reeves Flores, 2010; John C. 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 come up with 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) famous 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 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, 2022) of stone artifacts. The Triple P framework recognizes that experimental archaeology can greatly benefit from diverse perspectives and contributions from multiple stakeholders. By engaging researchers, practitioners, and local communities from different geographical locations, the framework acknowledges the importance of including voices from various cultural backgrounds and contexts (Pargeter et al., 2023: 164). This emphasis on collaboration and inclusivity allows for a more nuanced understanding of the complexities of raw material procurement (Batalla, 2016), selection (Arthur, 2021), pre-treatment (Maloney & Street, 2020), production (Griffin et al., 2013), and use (Martellotta et al., 2022) across different regions. Furthermore, the Triple P framework promotes the recognition and value of local knowledge and expertise. It acknowledges that communities living in specific geographical areas possess unique insights and understanding of their cultural heritage. By involving these local communities 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).

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

As implied in its name, 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 that partially contributed to the discipline-wide paradigm shift in the 1980s (Lake, 2014: 264-265). Equifinality refers to the phenomenon where 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 behavioral processes and intentions 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 at these three levels and thereby evaluate the probability of certain behavioral mechanisms behind a given archaeological assemblage (Reynolds, 1999; Stout & Hecht, 2023). The importance of specifying and documenting the context information of both the experiment as well as the phenomenon of interest has also been recently highlighted in psychological sciences (Holleman et al., 2020).

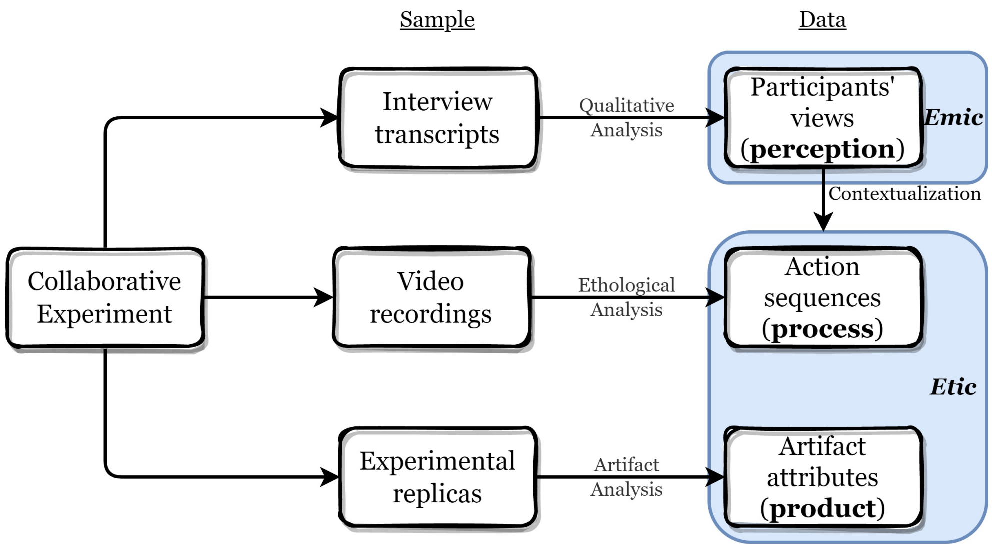

Figure 1: A schematic diagram demonstrating how to operationalize the Perception-Process-Product conceptual framework.

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, or photos and illustrations, or high-resolution 3D scans of individual artifact 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 it should be emphasized that 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. Besides, adopting good habits in spreadsheet data organization is strongly recommended (Broman & Woo, 2018).

4.2 Process-level data

While formal ethological methods that are widely used in the description and analysis of non-human animal behavior (Fragaszy & Mangalam, 2018) still largely fall into oblivion among archaeologists, the attempts of reconstructing behavioral sequences involved in the manufacture of material remains are not infrequent. One such example is cognigram, which was first systematically developed and applied in the archaeological research by Haidle (Miriam N. Haidle, 2009, 2010; Miriam Noël Haidle, 2023; Lombard & Haidle, 2012). Cognigram essentially represents an abstracting process of a series of action sequences achieving a similar goal. This approach is a powerful and elegant yet limited by the curse of expertise (Hinds, 1999), 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 assumes clear causal thinking between each step in an idealistic manner. However, this may be biased given that 1) experiments and modeling studies in cognitive science have shown that novices often feature a low planning depth (Opheusden et al., 2023) and 2) ethnographic studies demonstrated that even expert practitioners in traditional societies can have a different set of perception on the causal structure of how certain behaviors will modify the raw materials (Harris et al., 2021).

Consequently, we need to accumulate more real-world data by recording a large amount videos of toolmaking and conducting systematical 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 degrees of coding complexity (Cueva-Temprana et al., 2019; Mahaney, 2014). 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 (Mobbs et al., 2021), etc.

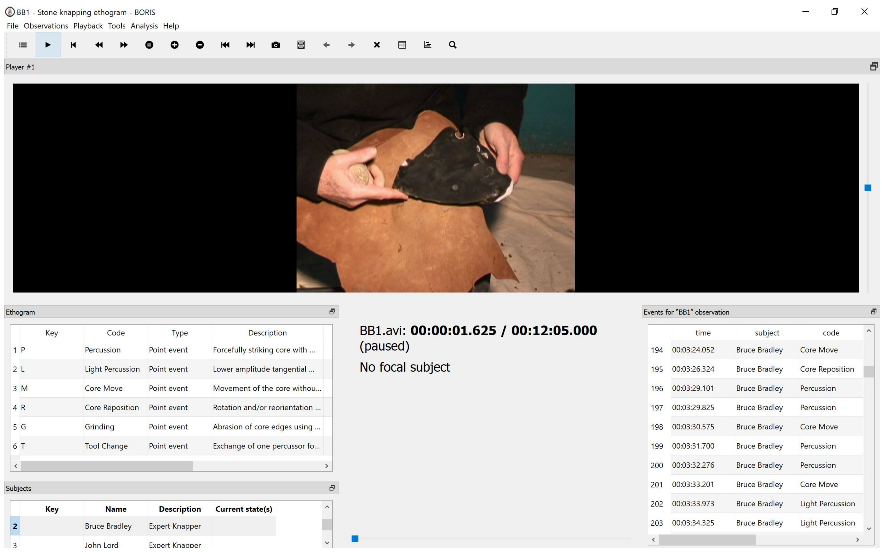

Figure 2: An example of coding Bruce Bradley's handaxe knapping session using the action grammer and BORIS software.

Table 1: A modiefied version of the original action grammer 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. Often occurs during core inspection
Core Reposition	Rotation and/or reorientation of the core involving repositioning of the hand. Often associated with the transition to a new percussion target
Grinding	Abrasion of core edges using a hammerstone
Tool Change	Exchange of one percussor for another

4.3 Perception-level data

Ethnographies revolving around general archaeological practices (Edgeworth, 2006), experimental archaeology as a field (Reeves Flores, 2012), as well as practices of specific technologies like flintknapping, including both WEIRD (John, C. Whittaker, 2004) and non-WEIRD populations(Arthur, 2018; Stout, 2002), are far from novel. However, it 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 that is otherwise irretrievable and thus should occupy a unique niche in experimental archaeology. This also echoes the post-positivist turn in psychology, a field that is known for the development of experimental methods, 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, ethnographers can capture the subtle nuances of perception, such as sensory experiences, social interactions, and cultural meanings associated with the experimental activities. 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 artifacts 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.

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 et al., 2017), which unfortunately still does not exist yet in experimental archaeology and likely won't be available in the near future. Among the three dimensions of the Tripe 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 highest 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 that was originally designed for developmental scientists, as the main data curation platform, where researchers can freely upload video files and related metadata that can connect with different types of data within the same project.

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 behavioral archaeology (emphasis on video recording of behavioral 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 a huge 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 would like to thank Dietrich Stout for his helpful comments on earlier drafts of this article. I am also grateful for my collaborators who help me turn this idea into reality: Russell Cutts (Oxford College of Emory University), Nada Khreisheh (The Ancient Technology Centre), Justin Pargeter (New York University), Ron Shimelmitz (University of Haifa), and particularly Mark Moore (University of New England). This study was supported by a research grant from the Leakey Foundation titled “Inferring skill reproduction from stone artifacts: A middle-range approach.”

References

Armstrong, S., & Mindermann, S. (2018). *Occam's razor is insufficient to infer the preferences of irrational agents*. *31*. https://proceedings.neurips.cc/paper/2018/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html

Arnaud, V., Pellegrino, F., Keenan, S., St-Gelais, X., Mathevon, N., Levréro, F., & Coupé, C. (2023).

Improving the workflow to crack Small, Unbalanced, Noisy, but Genuine (SUNG) datasets in bioacoustics: The case of bonobo calls. *PLOS Computational Biology*, *19*(4), e1010325. https://doi.org/10.1371/journal.pcbi.1010325

Arthur, K. W. (2018). *The lives of stone tools: Crafting the status, skill, and identity of flintknappers* (1st edition). 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

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

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

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*(5), 1232–1254. https://doi.org/10.1086/261651

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

Cartwright, N. (2007). Are RCTs the Gold Standard? *BioSocieties*, *2*(1), 11–20. https://doi.org/10.1017/S1745855207005029

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

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.

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

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

Edgeworth, M. (Ed.). (2006). *Ethnographies of archaeological practice: Cultural encounters, material transformations*. AltaMira Press.

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

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

Fletcher, B., Gheorghe, A., Moore, D., Wilson, S., & Damery, S. (2012). Improving the recruitment activity of clinicians in randomised controlled trials: a systematic review. *BMJ Open*, *2*(1), e000496. https://doi.org/10.1136/bmjopen-2011-000496

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

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

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, Miriam 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, Miriam 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, Miriam Noël. (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.

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.

Ingersoll, D., & MacDonald, W. (1977). *Introduction* (D. Ingersoll, J. E. Yellen, & W. MacDonald, Eds.; p. xixvii). Columbia University Press.

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

Lake, M. W. (2014). Trends in Archaeological Simulation. *Journal of Archaeological Method and Theory*, *21*(2), 258–287. https://doi.org/10.1007/s10816-013-9188-1

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

Lombard, M., & Haidle, M. N. (2012). Thinking a Bow-and-arrow Set: Cognitive Implications of Middle Stone Age Bow and Stone-tipped Arrow Technology. *Cambridge Archaeological Journal*, *22*(2), 237–264. https://doi.org/10.1017/S095977431200025X

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., d'Alpoim Guedes, J., Barton, C. M., Bates, L. A., Baxter, M., Bevan, A., Bollwerk, E. A., Bocinsky, R. K., Brughmans, T., Carter, A. K., Conrad, C., Contreras, D. A., Costa, S., Crema, E. R., Daggett, A., Davies, B., Drake, B. L., Dye, T. S., France, P., … Wren, C. D. (2017). Open science in archaeology. *SAA Archaeological Record*, *17*(4), 8–14. http://onlinedigeditions.com/publication/?i=440506

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

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

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* (H. G. Nami, Ed.; pp. 91–168). Ediciones de ArqueologÌa Contempor·nea.

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

Nichols, A. (2007). Causal Inference with Observational Data. *The Stata Journal*, *7*(4), 507–541. https://doi.org/10.1177/1536867X0800700403

Opheusden, B. van, Kuperwajs, I., Galbiati, G., Bnaya, Z., Li, Y., & Ma, W. J. (2023). Expertise increases planning depth in human gameplay. *Nature*, 1–6. https://doi.org/10.1038/s41586-023-06124-2

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.

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. (2010). *Creating a history of experimental archaeology* (D. Millson, Ed.; pp. 29–45). Oxbow Books.

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* (A. Harding, Ed.; 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

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, 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

Schulz, K. F., & Grimes, D. A. (2002). Allocation concealment in randomised trials: defending against deciphering. *The Lancet*, *359*(9306), 614–618. https://doi.org/10.1016/S0140-6736(02)07750-4

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

Speich, B., Schur, N., Gryaznov, D., Niederhäusern, B. von, Hemkens, L. G., Schandelmaier, S., Amstutz, A., Kasenda, B., Pauli-Magnus, C., Ojeda-Ruiz, E., Tomonaga, Y., McCord, K., Nordmann, A., Elm, E. von, Briel, M., Schwenkglenks, M., & Groups, a. collaboration of the M. (MAking. R. T. A. and A. (Adherence. to S. P. I. Re. for interventional trials). S. (2019). Resource use, costs, and approval times for planning and preparing a randomized clinical trial before and after the implementation of the new Swiss human research legislation. *PLOS ONE*, *14*(1), e0210669. https://doi.org/10.1371/journal.pone.0210669

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

Stout, D., & Hecht, E. (2023). *Evolutionary neuroarchaeology* (T. Wynn, K. A. Overmann, & F. L. Coolidge, Eds.; p. C14.S1C14.S11). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780192895950.013.14

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. (2022). A collaborative model for lithic shape digitization in museum settings. *Lithic Technology*, *0*(0), 1–12. 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, John C. (1994). *Flintknapping: Making and Understanding Stone Tools*. University of Texas Press.

Whittaker, John, C. (2004). *American Flintknappers: Stone Age Art in the Age of Computers*. University of Texas Press.

Whittaker, John C., & Stafford, M. (1999). Replicas, fakes, and art: The twentieth century stone age and its effects on archaeology. *American Antiquity*, *64*(2), 203–214. https://doi.org/10.2307/2694274

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