SAA
SOCIETY FOR AMERICAN ARCHAEOLOGY

ARTICLE

Variation Matters: Expanding the Scope of Experimental Archaeology

Cheng Liu (iD)

Department of Anthropology, Emory University, Atlanta, GA, USA
Email: cheng.liu@emory.edu

Abstract

This article aims to expand the scope of experimental archaeology to emphasize multilevel variation and interactions across the levels of perception, actions, and outcomes. Such an approach, loosely formulated as the Perception-Process-Product ("Triple P") framework, offers 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 nontraditional research participants, such as hobbyists and novices; and (3) adopting a workflow that normalizes the collection and curation of ethological and ethnographic data in experimental projects. Serving as a heuristic device, this alternative mode of knowledge production is highly flexible in nature, where each single component is detachable as dictated by individual research questions.

Resumen

El objetivo de este artículo es ampliar el alcance de la arqueología experimental para centrar la atención en la variación multinivel y las interacciones entre los rangos de percepción, acciones y resultados. Este enfoque, formulado vagamente como el marco Percepción-Proceso-Producto ("Triple P"), ofrece una explicación más fundamentada y rica del registro arqueológico del pasado. Consta de tres principios (1) reconocer la compensación inherente entre control y generalizabilidad en el diseño de la investigación experimental; (2) fomentar proyectos de colaboración que impliquen a participantes en la investigación de procedencias distintas no limitados a los tradicionales participantes, como aficionados y novatos; (3) adoptar un flujo de trabajo que normalice la recogida y conservación de datos etológicos y etnográficos en proyectos experimentales. Como dispositivo heurístico, este modo alternativo de producción de conocimientos es muy flexible por naturaleza, y cada uno de sus componentes es desechable o modificable en función de las preguntas de investigación individuales.

Keywords: experimental archaeology; ethological analysis; ethnographic research; curse of knowledge; collaborative knowledge production

Palabras clave: arqueología experimental; análisis etológico; investigación etnográfica; maldición del conocimiento; producción colaborativa de conocimientos

This article 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 (e.g., Blessing and Schmidt 2021; Domínguez-Rodrigo 2008; Reeves et al. 2009; Schmidt et al. 2019), whether explicitly or implicitly. This assumption acts to center inquiry around the reverse engineering of a past technology in a minimal or least-effort manner while ignoring the rich contextual information experimentation affords. When applied to the experimental study of ancient craftsmanship, Occam's razor—or the law of parsimony—implies that a technological solution that is simpler to reproduce is more likely to be the one used in the archaeological context. This is insufficient to infer the preferences of "irrational" agents possessing incomplete information (Mindermann and 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 optimality) that has been rejected by many anthropologists (Apicella et al. 2020; Henrich et al. 2001). Heyes (2012) similarly questioned the abuse of parsimony in animal behavioral research and proposed that new observational and experimental studies that allow testing of differential predictions become necessary when both a simple and a complex mechanism can explain the phenomenon of interest.

In fact, there are several reasons why past technologies may violate "parsimonious" assumptions of minimal manufacture complexity and optimal functional efficiency. In the evolution of technology, it is rather common that opaque causal perception and its resulting tendency of overimitation can lead to the widespread and long-lasting reproduction of technological solutions that are neither minimal in manufacture complexity nor optimal in functional efficiency. Overimitation 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 when compared with nonhuman primates, including chimpanzees (Horner and Whiten 2005), bonobos (Clay and Tennie 2018), and orangutans (Nielsen and Susianto 2010). Subsequent research further suggested that within human societies overimitation has been commonly observed among children across various cultural contexts (Berl and Hewlett 2015; Nielsen and Tomaselli 2010; Nielsen et al. 2014; Stengelin et al. 2020; Subiaul et al. 2016). Gergely and Csibra (2006) introduced "Sylvia's Recipe," which 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 having observed her mother throughout her childhood, who cut both ends of a ham. Later in life, her mother happened to watch her cooking, during which she noticed and asked Sylvia about the purpose of this step of the preparation. When Sylvia could not answer the question, her mother said that it was because she had not had a pan that was large enough to cook a full-sized ham. The commonality of this opaque causal perception has also been demonstrated in a recent study of Hadza bowmakers. Harris et alia (2021) found that even experienced bowmakers only possess limited causal knowledge regarding the design and construction of bows according to modern engineering principles—that is, 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 or simplification of manufacturing procedures. In this case, people are implicitly or explicitly aware of the existence of a more efficient solution, but they 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 even though such constraint no longer exists 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 and Baumhoff 1982; Mesoudi and O'Brien 2008).

Building on this critique of *Homo economicus* and the four strategies[1] of behavioral archaeology (Reid et al. 1975; Schiffer 2010), here I propose the Triple P framework, which aims to (1) amplify the expression of variation in experimental replicas (product) and their associated behavioral channels (process) as well as sensory experiences (perception) by experiments in diverse contexts and (2) 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: (a) acknowledging the inherent trade-off between control and generalizability in the experimental research design, (b) encouraging collaborative projects that involve geographically diverse and nontraditional research participants such as hobbyists and novices, and (c) adopting a workflow that normalizes the collection and curation of ethological and ethnographic data

in experimental projects. These two principles are developed to advocate for a pluralistic approach to the explanation of complex variation, which has received more attention in evolutionary anthropology (Antón and Kuzawa 2017) and 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 and Fryer 2023). Regarding the third principle of the Triple P framework, 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 tool kit available can sometimes inspire researchers to ask questions that are bold and transformative (Schmidt and Marwick 2020). Here, I will leverage the extensive corpus in experimental designs and inferences revolving around stone artifacts to demonstrate the necessity and potential of this framework.

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 (Degtiar and Rose 2023; Roe and Just 2009:1266–1267).[2] 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 semicontrolled 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 and Mendelsohn 2019; Sonkusare et al. 2019; Yarkoni 2022). In contrast, the past decades have witnessed experimental archaeology's growing research interests focusing on the robust inference of causal mechanisms while compromising generalizability in the explanation of material culture variation (Eren and Meltzer 2024; Eren et al. 2016; Lin et al. 2018; Marreiros et al. 2020). 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 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 prioritizes the material science aspect over the social science aspect of experimental archaeology. Similarly, standardized artificial materials such as 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, producing results with limited generalizability (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).

In contrast, less-controlled experiments, which have been traditionally known as naturalistic or actualistic experiments (for detailed terminological critiques, see Conrad et al. 2023; Eren and 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). With regard to the cases of cultural transmission experiments, a less-controlled experiment would involve the use of (1) natural rocks with varied morphology instead of standardized artificial materials and (2) 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 (Almaatouq et al. 2024), less-controlled experiments are designed to produce variation and their interactions. This feature is crucial and cannot be simply replaced by ethnographic records or ethnoarchaeology, because many Paleolithic technological components do not have analogs in contemporary nonindustrial societies (e.g., Arthur 2018; Stout

2002). Although 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 and Robins 2023). Despite the fact that one should not interpret any experiment as a direct representation of an actual past event (Eren and Meltzer 2024), less-controlled experiments can serve a heuristic role in hypothesis generation, aligning with the perspective of Lin et alia (2018:680–681), who proposed that the interaction between less-controlled and strictly controlled experiments "operates in a cyclical form of induction and deduction."

Many Places, Many Voices

Traditional 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 the dual identity of also being a 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—when communicating with others (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 such as wrist and elbow), preferences of percussor and raw material types, and 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 a single knapper, have been constantly framed as generalizations 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 nineteenth century (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 counterintuitive 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 such as shaving and cutting hard materials, which challenged the traditional perspective on producing sharp lateral edges as the sole purpose of stone toolmaking and shed 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 flintknapping 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. Several researchers have also pointed out that literature-informed archaeologists sometimes get lost in reconstructing previous archaeologists' reconstructions instead of searching for diverse solutions to better understand the actual archaeological phenomenon (Bell 2014; Currie 2022), which is another reason why we need the presence of hobbyists and novices in the community of experimental archaeology.

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 and Stout 2023; Ranhorn et al. 2020) and museum collection studies (Timbrell 2023) of stone artifacts. Furthermore, the Triple P framework acknowledges that communities living in specific geographical areas possess unique insights and understanding of their cultural

heritage (Arthur et al. 2024). This emphasis on team efforts and inclusivity allows for a more complete understanding of the nonutilitarian or unexpected aspects of raw material procurement (Batalla 2016) and selection (Arthur 2021), pretreatment (Maloney and 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 nonindustrial 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 and 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 articles 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 (Ouzman 2023). Moreover, considering the checkered disciplinary history of archaeology/anthropology featuring colonial exploitation, the changes in the evaluation system alone are not enough. This further highlights the need for adopting a community-based approach to fundamentally transform the power dynamics in archaeological knowledge production and distribution (Atalay 2012; La Salle 2010; Schneider and Hayes 2020).

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 (Eren et al. 2024; 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 different paths, whereas multifinality emerges when a similar process can lead to multiple ends. Although we cannot fully solve these two problems and accurately reconstruct the past behavioral 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
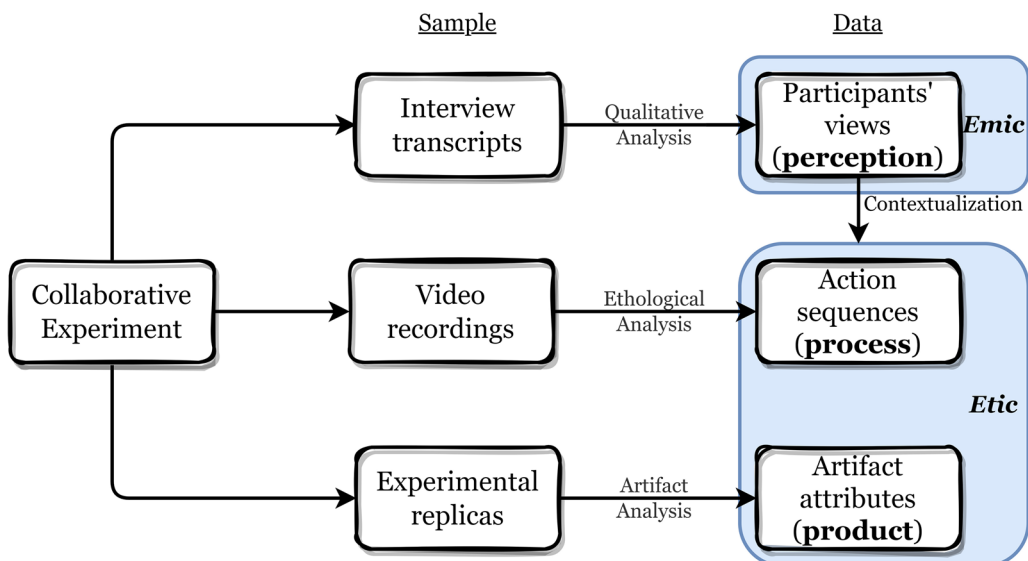

Figure 1. A schematic diagram demonstrating how to operationalize the Perception-Process-Product conceptual framework.

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).

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 artifacts 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 metadata. I also strongly recommend adopting good habits in spreadsheet data organization (Broman and Woo 2018).

Process-Level Data

Although systematic behavioral coding methods widely used in the study of nonhuman animal behavior (Fragaszy and Mangalam 2018) are still largely neglected among archaeologists, attempts to reconstruct behavioral sequences involved in the manufacture of material remains are not infrequent, ranging from the well-established *chaîne opératoire* approach (Audouze et al. 2017; Delage 2017; Dobres 1999; Porqueddu et al. 2023; Soressi and Geneste 2011) to the more recent cognigram method. To illustrate the benefits and drawbacks of existing analytical frameworks, I will use the cognigram as an example. First systematically developed and applied in archaeological research by Haidle (2009, 2010, 2014, 2023), a cognigram is a graphical representation of the reconstructed behavior behind archaeological artifacts 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, but it 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 understanding. However, this may be biased because (1) certain causal insights in stone fracture mechanics remained opaque to academic knappers until they were revealed through controlled experiments by Dibble and his colleagues (Li et al. 2022) and (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 tool-making videos and conducting systematic ethogram analysis. With the emergence of new software platforms such as BORIS (Friard and 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 alia (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 and 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 the measurement of the complexity of various technological systems—a classical topic in Paleolithic archaeology and the evolution of human cognition (Muller et al. 2017; Perreault et al. 2013). Unlike the traditional approaches resorting to the extraction and comparison of lithic attributes, Stout and colleagues (2021) recorded the videos of expert flintknappers reproducing Oldowan and Acheulean technologies and then manually parsed their knapping activities using action grammar, generating multiple sequences of actions. Borrowing tools from computational linguistics, they then calculated the transition probability between each action category across two technological systems, which provided an objective and quantifiable proxy for measuring technological complexity. Another scenario of its application is the measurement of behavioral similarity across individuals (Cristino et al. 2010; Mobbs et al. 2021), which is particularly relevant in the abovementioned cultural transmission experiments. Again,
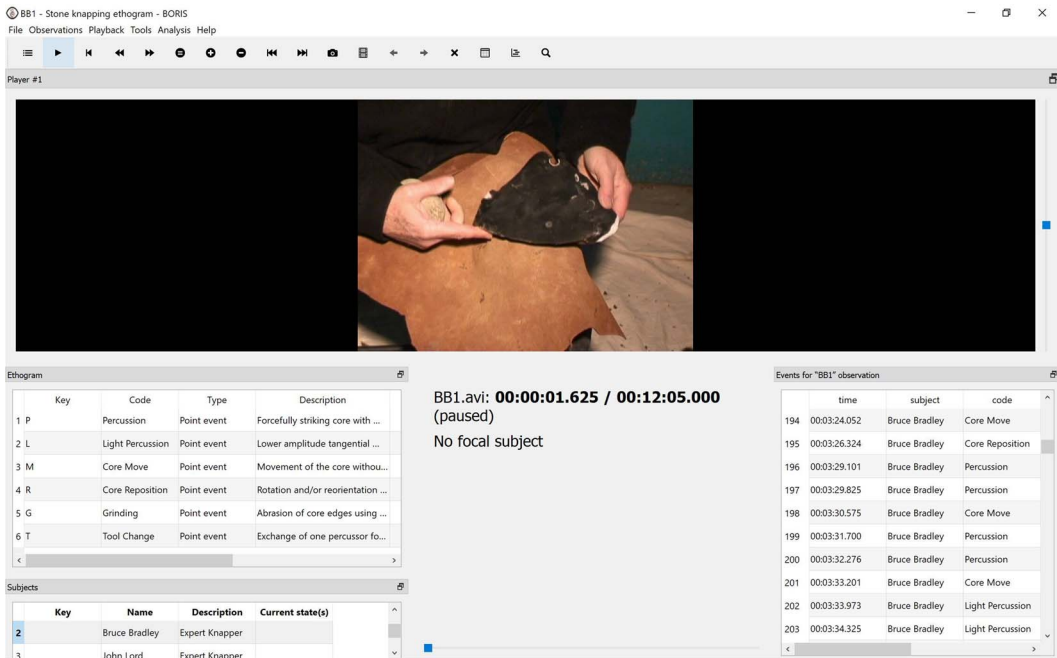
Figure 2. An example of coding a handaxe knapping session using the BORIS software.

given that the existing works on this topic focus mainly on the sole analysis of experimental replicas, many aspects of knapping skill-learning processes remain unclear. For example, how do different individual learning strategies (high-fidelity action copying vs. predominantly trial-and-error learning) affect the morphological variation of their final products? Or will learning behavioral conformity within a community of practice necessarily lead to homogeneity in the formation of lithic assemblages? To answer these questions, the quantitative analysis of process-level data associated with the product-level data becomes necessary. Behatrix (https://www.boris.unito.it/behatrix/; Figure 3), a sister software of BORIS, allows us to calculate the action sequence (dis)similarity between novice learners and expert demonstrators / fellow novice learners using established algorithms (for an application of analyzing play behavior similarity among gorillas, see Cordoni et al. 2022).

Table 1. A Modified Version (Liu et al. 2024) of the Original Action Grammar Presented in Stout and Colleagues (2021).

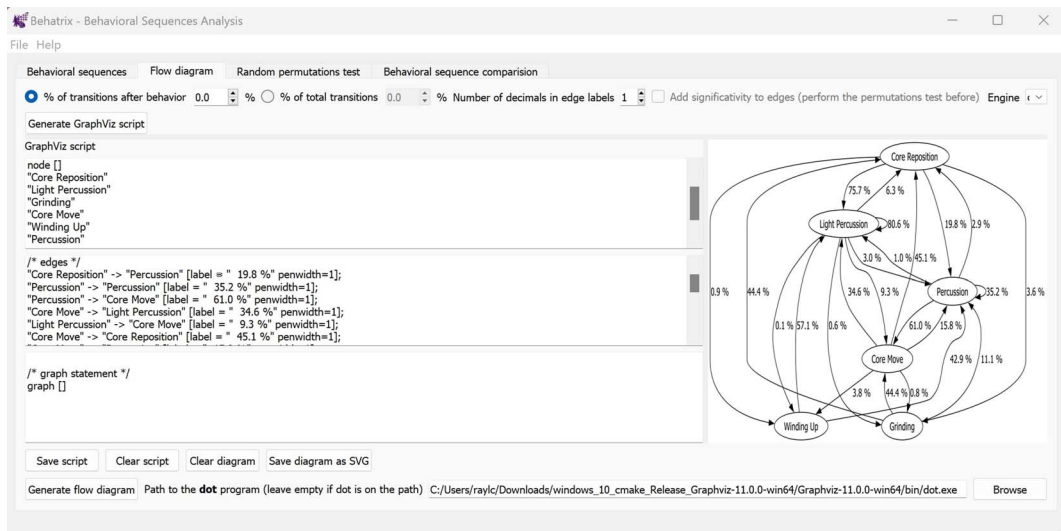| Action	Definition
Percussion	Forcefully striking core with percussor (hammerstone or antler billet) in such a way as to potentially remove a flake. Each strike should be counted as a single action.
Light Percussion	Lower-amplitude tangential strike to the tool edge of the kind often employed for platform preparation. Each strike should be counted as a single action.
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 toward the core that do not lead to the detachment of flakes, which can either be in direct contact with cores or not.

Figure 3. The user interface of Behatrix displaying the transition probability between each action category in a handaxe knapping session.

Perception-Level Data

Direct applications of ethnographic research in experimental archaeology as a field (Reeves Flores 2012) and practices of specific technologies such as flintknapping—including contemporary US hobbyists (Whittaker 2004) and knapping practitioners in various nonindustrial societies (Arthur 2018; Stout 2002)—are far from novel. However, ethnography has never been formally recognized as a legitimate research method in mainstream experimental archaeology. Echoing 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. Consequently, it should occupy a unique niche in experimental archaeology. Within the broader context of burgeoning interest in mixed-method research in contemporary social science (Creswell and 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 and 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 and Will 2021; Roepstorff 2008), sensory experiences (Day 2013; O'Neill and O'Sullivan 2019; Skeates and Day 2019), 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 both the research question and ad hoc interaction. One potential application of ethnographic methods in the experimental archaeology of stone artifacts is asking knappers about the intentions of each action and seeing how it matches with the results as revealed by lithic analysis of replicas, which can provide crucial contextual information that addresses the issues of equifinality and multifinality in the formation of lithic assemblage. For example, serial formation of step fracture on the debitage surface is commonly interpreted as unintentional mistakes indicative of novice knappers, whereas in some cases, researchers treat it as evidence of deliberate core rejuvenation (Akerman 1993:126). The accumulation of testimonies by participating knappers in terms of their intended outcome becomes useful in this scenario, although these materials should be examined in combination with the relevant product- and process-level data in a careful manner. 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—that is, "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 becomes necessary.

Multilevel Sample and 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 and Birch 2018), which unfortunately still does not exist in experimental archaeology. The accessibility and availability of experimental data can foster collaboration and enhance the reproducibility and transparency of research findings, because others can verify and validate the results by examining the original data. Moreover, a centralized database also promotes data preservation and long-term accessibility. Storing experimental data in a structured and organized manner safeguards valuable information from potential loss or degradation over time. This preservation ensures that the data remain accessible for future researchers, avoiding the loss of valuable insights and preventing the need for unnecessary 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; the perception-level data formats mainly include audio files and their transcribed texts; and videos are the main vector of process-level data, a rather nontraditional data format in archaeological research featuring the largest file size compared with the other two. Consequently, following data-sharing principles of FAIR (Nicholson et al. 2023; Wilkinson et al. 2016) and CARE (Carroll et al. 2020; Gupta et al. 2023), 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: (1) no cost from the side of researchers, (2) long-term data security monitored by a specialized maintenance team, and (3) the fostering of potential collaborations between experimental archaeologists and developmental psychologists.

On top of the digital data curation, an easily ignored but crucial aspect regarding the integrity and reliability of research in experimental archaeology is the long-term and proper curation of physical specimens produced during experiments, which is particularly relevant to the product-level data. Haythorn et alia (2018) sharply pointed out that rerunning statistical analyses on a publicly available spreadsheet containing incorrect lithic projectile attribute data would be meaningless. In this case, a reexamination of the actual experiment samples becomes necessary. Moreover, it is likely that new research questions can only be answered through direct observation and measurement of the experimental assemblages themselves instead of data readily available in previously compiled spreadsheets (Eren et al. 2016). The existence of these possibilities therefore requires that an experimental assemblage of interest should be curated in public institutions with easy access and detailed contextual information (Haythorn et al. 2018).

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 postprocessual 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. It should be noted that this alternative mode of knowledge production is not necessarily a bundle sale, where each single component is independent and detachable according to the individual research question. Instead, it can serve as a heuristic tool to inspire potential readers to explore a broader range of data collection and analysis strategies. To summarize, the innovativeness, flexibility, and inclusiveness of the Triple P conceptual framework have 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.

Notes

1. Strategy 2 is particularly relevant here, which is "the pursuit of general questions in present-day material culture in order to acquire laws for making useful behavioral inferences" (Schiffer 2010:6).
2. It should be noted that the definitions of internal and external validity constructs here solely concern the interpretation of experimental design and therefore are different from the one previously proposed by Lycett and Eren (Eren et al. 2016; Lycett and Eren 2013).

References Cited

Acerbi, Alberto, Magnus Enquist, and Stefano Ghirlanda. 2009. Cultural Evolution and Individual Developm0ent of Openness and Conservatism. *PNAS* 106(45):18931–18935. https://doi.org/10.1073/pnas.0908889106.

Akerman, Kim. 1993. The Status of the Horsehoof Core. *Records of the Australian Museum*, *Supplement* 17:125–127.

Almaatouq, Abdullah, Thomas L. Griffiths, Jordan W. Suchow, Mark E. Whiting, James Evans, and Duncan J. Watts. 2024. Beyond Playing 20 Questions with Nature: Integrative Experiment Design in the Social and Behavioral Sciences. *Behavioral and Brain Sciences* 47. https://doi.org/10.1017/S0140525X22002874.

Antón, Susan C., and Christopher W. Kuzawa. 2017. Early Homo, Plasticity and the Extended Evolutionary Synthesis. *Interface Focus* 7(5). https://doi.org/10.1098/rsfs.2017.0004.

Apicella, Coren, Ara Norenzayan, and Joseph Henrich. 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, Kathryn Weedman. 2018. *The Lives of Stone Tools: Crafting the Status, Skill, and Identity of Flintknappers*. University of Arizona Press, Tucson.

Arthur, Kathryn Weedman. 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.

Arthur, Kathryn Weedman, Ran Barkai, Catherine Allen, Ella Assaf Shpayer, Bar Efrati, Meir Finkel, Dov Ganchrow, et al. 2024. Ancestral Stones and Stone Stories: Reimagining Human Relationships with Stone from the Paleolithic to the Present. *Archaeologies* 20(1):1–23. https://doi.org/10.1007/s11759-024-09502-y.

Atalay, Sonya. 2012. *Community-Based Archaeology: Research with, by, and for Indigenous and Local Communities*. University of California Press, Berkeley.

Audouze, Françoise, Pierre Bodu, Claudine Karlin, Michèle Julien, Jacques Pelegrin, and Catherine Perlès. 2017. Leroi-Gourhan and the Chaîne Opératoire: A Response to Delage. *World Archaeology* 49(5):718-723. https://doi.org/10.1080/00438243.2017.1416012.

Barrett, H. Clark. 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, Arlys Nicolás. 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.

Bell, Martin. 2014. Experimental Archaeology at the Crossroads: A Contribution to Interpretation or Evidence of "Xeroxing"? In *Material Evidence: Learning from Archaeological Practice*, edited by Robert Chapman and Alison Wylie, pp. 42–58. Routledge, London.

Berl, Richard E. W., and Barry S. Hewlett. 2015. Cultural Variation in the Use of Overimitation by the Aka and Ngandu of the Congo Basin. *PLoS ONE* 10(3):e0120180. https://doi.org/10.1371/journal.pone.0120180.

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

Blessing, Matthias A., and Patrick Schmidt. 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, Karl W., and Kara H. Woo. 2018. Data Organization in Spreadsheets. *American Statistician* 72(1):2–10. https://doi.org/10.1080/00031305.2017.1375989.

Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, et al. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19(1):43. https://doi.org/10.5334/dsj-2020-043.

Clay, Zanna, and Claudio Tennie. 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.

Conrad, Grace, Savannah Hough, Jacob Baldino, Nicholas Gala, Briggs Buchanan, Robert S. Walker, Alastair Key, Brian G. Redmond, Michelle R. Bebber, and Metin I. Eren. 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.

Cordoni, Giada, Luca Pirarba, Stéphanie Elies, Elisa Demuru, Jean-Pascal Guéry, and Ivan Norscia. 2022. Adult–Adult Play in Captive Lowland Gorillas (*Gorilla gorilla gorilla*). *Primates* 63(3):225–235. https://doi.org/10.1007/s10329-022-00973-7.

Crabtree, Don E. 1977. The Obtuse Angle as a Functional Edge. In *Experimental Archaeology*, edited by Daniel Ingersoll, John E. Yellen, and William MacDonald, pp. 38–51. Columbia University Press, New York.

Creswell, John W., and Vicki L. Plano Clark. 2017. *Designing and Conducting Mixed Methods Research*. 3rd ed. Sage Publications, Los Angeles.

Cristino, Filipe, Sebastiaan Mathôt, Jan Theeuwes, and Iain D. Gilchrist. 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, Arturo, Diego Lombao, Juan Ignacio Morales, Nuria Geribàs, and Marina Mosquera. 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, Scott. 2021. *Causal Inference: The Mixtape*. Yale University Press, New Haven, Connecticut. https://doi.org/10.2307/j.ctv1c29t27.

Currie, Adrian. 2022. Speculation Made Material: Experimental Archaeology and Maker's Knowledge. *Philosophy of Science* 89(2):337–359. https://doi.org/10.1017/psa.2021.31.

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

Day, Jo. 2013. *Making Senses of the Past: Toward a Sensory Archaeology*. Southern Illinois University Press, Carbondale.

Degtiar, Irina, and Sherri Rose. 2023. A Review of Generalizability and Transportability. *Annual Review of Statistics and Its Application* 10:501–524. https://doi.org/10.1146/annurev-statistics-042522-103837.

Delage, Christophe. 2017. "Once upon a Time . . . the (Hi)story of the Concept of the *Chaîne Opératoire* in French Prehistory. *World Archaeology* 49(2):158–173. https://doi.org/10.1080/00438243.2017.1300104.

Dobres, Marcia-Anne. 1999. Technology's Links and Chaînes: The Processual Unfolding of Technique and Technician. In *The Social Dynamics of Technology: Practice, Politics, and World Views*, edited by Dobres Marcia-Anne and Christopher R. Hoffman, pp. 124–146. Smithsonian Institution, Washington, DC.

Domínguez-Rodrigo, Manuel. 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, Kristina. 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, Metin I., Michelle R. Bebber, Lawrence Mukusha, Michael Wilson, Andrew R. Boehm, Briggs Buchanan, G. Logan Miller, et al. 2024. Experimental Bison Butchery Using Replica Hafted Clovis Fluted Points and Large Handheld Flakes. *Journal of Archaeological Science: Reports* 55. https://doi.org/10.1016/j.jasrep.2024.104480.

Eren, Metin I., Stephen J. Lycett, Robert J. Patten, Briggs Buchanan, Justin Pargeter, and Michael J. O'Brien. 2016. Test, Model, and Method Validation: The Role of Experimental Stone Artifact Replication in Hypothesis-Driven Archaeology. *Ethnoarchaeology* 8(2):103–136. https://doi.org/10.1080/19442890.2016.1213972.

Eren, Metin I., and David J. Meltzer. 2024. Controls, Conceits, and Aiming for Robust Inferences in Experimental Archaeology. *Journal of Archaeological Science: Reports* 53. https://doi.org/10.1016/j.jasrep.2024.104411.

Faniel, Ixchel M., Anne Austin, Eric Kansa, Sarah Whitcher Kansa, Phoebe France, Jennifer Jacobs, Ran Boytner, and Elizabeth Yakel. 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.

Fragaszy, Dorothy M., and Madhur Mangalam. 2018. Chapter Five - Tooling. In *Advances in the Study of Behavior (Vol. 50)*, edited by Marc Naguib, Louise Barrett, Susan D. Healy, Jeffrey Podos, Leigh W. Simmons, and Marlene Zuk, pp. 177–241. Academic Press, Cambridge, Massachusetts. https://doi.org/10.1016/bs.asb.2018.01.001.

Friard, Olivier, and Marco Gamba. 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, György, and Gergely Csibra. 2006. Sylvia's Recipe: The Role of Imitation and Pedagogy in the Transmission of Cultural Knowledge. In *Roots of Human Sociality: Culture, Cognition and Interaction*, edited by Stephen C. Levinson and Nicholas J. Enfield, pp. 229–255. Berg, Oxford.

Ghirlanda, Stefano, Magnus Enquist, and Mayuko Nakamaru. 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, Rick, Karen Adolph, David Millman, Lisa Steiger, and Dylan Simon. 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, Geoffrey. 2019. The Sociality of Enskilment. *Ethnos* 84(3):508–524. https://doi.org/10.1080/00141844.2018.1455726.

Griffin, Darren, Delta Lucille Freedman, Bill Nicholson, Fiona McConachie, and Alexander Parmington. 2013. The *Koorong* Project: Experimental Archaeology and Wurundjeri Continuation of Cultural Practices. In *Excavations, Surveys and Heritage Management in Victoria Volume 2*, edited by Susan Lawrence, David Frankel, Caroline Spry, Shaun Cunning, and Ilya Berelo, pp. 59–65. La Trobe University Archaeology Program, Bundoora, Australia.

Gupta, Neha, Andrew Martindale, Kisha Supernant, and Michael Elvidge. 2023. The CARE Principles and the Reuse, Sharing, and Curation of Indigenous Data in Canadian Archaeology. *Advances in Archaeological Practice* 11(1):76–89. https://doi.org/10.1017/aap.2022.33.

Haidle, Miriam N. 2009. How to Think a Simple Spear. In *Cognitive Archaeology and Human Evolution*, edited by Sophie A. de Beaune, Frederick L. Coolidge, and Thomas Wynn, pp. 57–73. Cambridge University Press, Cambridge.

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):S149–166. https://doi.org/10.1086/650295.

Haidle, Miriam N. 2014. Building a Bridgean Archeologist's Perspective on the Evolution of Causal Cognition. *Frontiers in Psychology* 5:1472. https://doi.org/10.3389/fpsyg.2014.01472.

Haidle, Miriam N. 2023. Cognigrams: Systematically Reconstructing Behavioral Architectures as a Basis for Cognitive Archaeology. In *Oxford Handbook of Cognitive Archaeology*, edited by Thomas Wynn, Karenleigh A. Overmann, and Frederick L. Coolidge, pp. C12S1–C12S8. Oxford University Press, Oxford. https://doi.org/10.1093/oxfordhb/9780192895950.013.12.

Harris, Jacob A., Robert Boyd, and Brian M. Wood. 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.

Haythorn, Richard, Briggs Buchanan, and Metin I. Eren. 2018. A New Look at Flaked Stone Projectiles from the Mixter Site (33-ER-4), Erie County, Ohio, USA. *Lithic Technology* 43(3):166–171. https://doi.org/10.1080/01977261.2018.1479950.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 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, Miguel A., and James M. Robins. 2023. *Causal Inference: What If*. CRC Press, Boca Raton, Florida.

Heyes, Cecilia. 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, Pamela 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, Peter. 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, Gijs A., Ignace T. C Hooge, Chantal Kemner, and Roy S. Hessels. 2020. The "Real-World Approach" and Its Problems: A Critique of the Term Ecological Validity. *Frontiers in Psychology* 11:721. https://doi.org/10.3389/fpsyg.2020.00721.

Horner, Victoria, and Andrew Whiten. 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, Jeremy. 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, Shumon T., and Manuel Will. 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. Lewis. 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, Mahdi, Elahe Daviran, and Mostafa Taqavi. 2022. The QWERTY Keyboard from the Perspective of the Collingridge Dilemma: Lessons for Co-Construction of Human-Technology. *AI & SOCIETY* 39:1229–1241. https://doi.org/10.1007/s00146-022-01573-1.

La Salle, Marina J. 2010. Community Collaboration and Other Good Intentions. *Archaeologies* 6(3):401–422. https://doi.org/10.1007/s11759-010-9150-8.

Li, Li, Sam C. Lin, Shannon P. McPherron, Aylar Abdolahzadeh, Annie Chan, Tamara Dogandžić, Radu Iovita, et al. 2022. A Synthesis of the Dibble et al. Controlled Experiments into the Mechanics of Lithic Production. *Journal of Archaeological Method and Theory* 30:1284–1325. https://doi.org/10.1007/s10816-022-09586-2.

Lin, Sam C., Zeljko Rezek, and Harold L. Dibble. 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, Cheng, Nada Khreisheh, Dietrich Stout, and Justin Pargeter. 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. https://doi.org/10.1016/j.jasrep.2023.103974.

Liu, Cheng, and Dietrich Stout. 2023. Inferring Cultural Reproduction from Lithic Data: A Critical Review. *Evolutionary Anthropology* 32(2):83–99. https://doi.org/10.1002/evan.21964.

Liu, Cheng, Xuyang Yan, Cheng Ding, Tianhao Zhao, Shaya Jannati, Cynthia Martinez, and Dietrich Stout. 2024. Human Stone Toolmaking Action Grammar (HSTAG): A Challenging Benchmark for Fine-grained Motor Behavior Recognition. In *2024 IEEE 11th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–8. IEEE, New York. https://doi.org/10.1109/DSAA61799.2024.10722814.

Lombao, D., M. Guardiola, and M. Mosquera. 2017. Teaching to Make Stone Tools: New Experimental Evidence Supporting a Technological Hypothesis for the Origins of Language. *Scientific Reports* 7(1). https://doi.org/10.1038/s41598-017-14322-y.

Lycett, Stephen J., and Metin I. Eren. 2013. Levallois Lessons: The Challenge of Integrating Mathematical Models, Quantitative Experiments and the Archaeological Record. *World Archaeology* 45(4):519–538. https://doi.org/10.1080/00438243.2013.821670.

Lyons, Derek E., Andrew G. Young, and Frank C. Keil. 2007. The Hidden Structure of Overimitation." *PNAS* 104(50): 19751–19756. https://doi.org/10.1073/pnas.0704452104.

Mahaney, Robert Allen. 2014. Exploring the Complexity and Structure of Acheulean Stoneknapping in Relation to Natural Language. *PaleoAnthropology* 2014:586–606.

Malafouris, Lambros. 2013. *How Things Shape the Mind: A Theory of Material Engagement*. MIT Press, Cambridge, Massachusetts.

Maloney, Tim Ryan, and Mervyn Street. 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.

Marreiros, João, Telmo Pereira, and Radu Iovita. 2020. Controlled Experiments in Lithic Technology and Function. *Archaeological and Anthropological Sciences* 12:Article 110. https://doi.org/10.1007/s12520-020-01059-5.

Marshall, Yvonne. 2002. What Is Community Archaeology? *World Archaeology* 34(2):211–219. https://doi.org/10.1080/0043824022000007062.

Martellotta, Eva Francesca, Yinika L. Perston, Paul Craft, Jayne Wilkins, and Michelle C. Langley. 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, Ben, and Suzanne E. Pilaar Birch. 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, Alex, and Michael J. O'Brien. 2008. The Cultural Transmission of Great Basin Projectile-Point Technology I: An Experimental Simulation. *American Antiquity* 73(1):3–28. https://doi.org/https://doi.org/10.1017/S0002731600041263.

Milks, Annemieke, Christian Hoggard, and Matt Pope. 2023. Reassessing the Interpretative Potential of Ethnographic Collections for Early Hunting Technologies. *Journal of Archaeological Method and Theory* 31:1129–1151. https://doi.org/10.1007/s10816-023-09635-4.

Mindermann, Soren, and Stuart Armstrong. 2018. Occam's Razor Is Insufficient to Infer the Preferences of Irrational Agents. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 5603–5614. Curran Associates, Red Hook, New York.

Mobbs, Dean, Toby Wise, Nanthia Suthana, Noah Guzmán, Nikolaus Kriegeskorte, and Joel Z. Leibo. 2021. Promises and Challenges of Human Computational Ethology. *Neuron* 109(14):2224–2238. https://doi.org/10.1016/j.neuron.2021.05.021.

Montgomery, Lindsay M., and Tiffany C. Fryer. 2023. The Future of Archaeology Is (Still) Community Collaboration. *Antiquity* 97(394):795–809. https://doi.org/10.15184/aqy.2023.98.

Moody, Bryony, Tom Dye, Keith May, Holly Wright, and Caitlin Buck. 2021. Digital Chronological Data Reuse in Archaeology: Three Case Studies with Varying Purposes and Perspectives. *Journal of Archaeological Science: Reports* 40, Part A. https://doi.org/10.1016/j.jasrep.2021.103188.

Moore, Mark 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, Olivier. 2022. Cultural Conservatism. *Journal of Cognition and Culture* 22(5):406–420.

Muller, Antoine, Chris Clarkson, and Ceri Shipton. 2017. Measuring Behavioural and Cognitive Complexity in Lithic Technology throughout Human Evolution. *Journal of Anthropological Archaeology* 48:166–180. https://doi.org/10.1016/j.jaa.2017.07.006.

Muller, Antoine, Ceri Shipton, and Chris Clarkson. 2023. The Proceduralization of Hominin Knapping Skill: Memorizing Different Lithic Technologies. *Cambridge Archaeological Journal* 33(4):655–672. https://doi.org/10.1017/S0959774323000070.

Nami, Hugo, G. 2010. Theoretical Reflections on Experimental Archaeology and Lithic Technology: Issues on Actualistic Stone Tools Analysis and Interpretation. In *Experiments and Interpretation of Traditional Technologies: Essays in Honor of Errett Callahan*, edited by Hugo G. Nami, pp. 91–168. Ediciones de Arqueología Contemporánea, Buenos Aires.

Nastase, Samuel A., Ariel Goldstein, and Uri Hasson. 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.

Nicholson, Christopher, Sarah Kansa, Neha Gupta, and Rachel Fernandez. 2023. Will It Ever Be FAIR? Making Archaeological Data Findable, Accessible, Interoperable, and Reusable. *Advances in Archaeological Practice* 11(1):63–75. https://doi.org/10.1017/aap.2022.40.

Nielsen, Mark, Ilana Mushin, Keyan Tomaselli, and Andrew Whiten. 2014. Where Culture Takes Hold: "Overimitation" and Its Flexible Deployment in Western, Aboriginal, and Bushmen Children. *Child Development* 85(6):2169–2184. https://doi.org/10.1111/cdev.12265.

Nielsen, Mark, and Erna W. E. Susianto. 2010. Failure to Find Over-Imitation in Captive Orangutans (Pongo pygmaeus): Implications for Our Understanding of Cross-Generation Information Transfer. In *Developmental Psychology*, edited by Johan Håkansson, pp. 153–167. Nova Science, New York.

Nielsen, Mark, and Keyan Tomaselli. 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.

O'Neill, Brendan, and Aidan O'Sullivan. 2019. Experimental Archaeology and (Re)-Experiencing the Senses of the Medieval World. In *The Routledge Handbook of Sensory Archaeology*, edited by Robin Skeates and Jo Day, pp. 451–466. Routledge, London.

Outram, Alan K. 2008. Introduction to Experimental Archaeology. *World Archaeology* 40(1):1–6.

Ouzman, Sven. 2023. Authorship, Attribution and Acknowledgment in Archaeology. *Australian Archaeology* 89(1):66–70. https://doi.org/10.1080/03122417.2023.2190497.

Pargeter, Justin, Cheng Liu, Megan Beney Kilgore, Aditi Majoe, and Dietrich Stout. 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.

Perreault, Charles, P. Jeffrey Brantingham, Steven L. Kuhn, Sarah Wurz, and Xing Gao. 2013. Measuring the Complexity of Lithic Technology. *Current Anthropology* 54(S8):S397–406. https://doi.org/10.1086/673264.

Pfleging, Johannes, Radu Iovita, and Jonas Buchli. 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.

Porqueddu, Marie-Elise, Claudia Sciuto, and Anaïs Lamesa. 2023. Reconsidering the Chaîne Opératoire: At the Crossroad between People and Materials. *Open Archaeology* 9(1). https://doi.org/10.1515/opar-2022-0296.

Premo, Luke S. 2010. Equifinality and Explanation: Thoughts on the Role of Agent-Based Modeling in Postpositivist Archaeology. In *Simulating Change: Archaeology into the Twenty-First Century*, edited by Andre Costopoulos and Mark W. Lake, pp. 28–37. University of Utah Press, Salt Lake City.

Proffitt, T., A. Bargalló, and I. de la Torre. 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, Tadeg, and Tamsin C. German. 2021. A Simple Definition of "sIntentionally." *Cognition* 214:104806. https://doi.org/10.1016/j.cognition.2021.104806.

Ranhorn, Kathryn L., Justin Pargeter, and L. S. Premo. 2020. Investigating the Evolution of Human Social Learning through Collaborative Experimental Archaeology. *Evolutionary Anthropology* 29(2):53–55. https://doi.org/10.1002/evan.21823.

Reeves, Dan, Rick Bury, and David W. Robinson. 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, Jodi. 2012. Experimental Archaeology: An Ethnography of Its Perceived Value and Impact in Archaeological Research. PhD dissertation, Department of Archaeology, University of Exeter, Exeter. https://ore.exeter.ac.uk/repository/handle/10871/9041.

Reid, J. Jefferson, Michael B. Schiffer, and William L. Rathje. 1975. Behavioral Archaeology: Four Strategies. *American Anthropologist* 77(4):864–869. https://doi.org/10.1525/aa.1975.77.4.02a00090.

Reynolds, Peter J. 1999. The Nature of Experiment in Archaeology. In *Experiment and Design: Archaeological Studies in Honour of John Coles*, edited by Anthony Harding, pp. 156–162. Oxbow Books, Oxford.

Roe, Brian E., and David R. Just. 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.

Roepstorff, Andreas. 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, Valentine, and Éva David. 2005. Planning Abilities as a Dynamic Perceptual-Motor Skill: An Actualist Study of Different Levels of Expertise Involved in Stone Knapping. In *Stone Knapping: The Necessary Preconditions for a Uniquely Hominin Behaviour*, edited by Valentine Roux and Blandine Brill, pp. 91–108. McDonald Institute for Archaeological Research, Cambridge.

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

Schiffer, Michael B. 2010. *Behavioral Archaeology: Principles and Practice*. Routledge, London.

Schillinger, Kerstin, Alex Mesoudi, and Stephen J. Lycett. 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, Patrick, Matthias Blessing, Maxime Rageot, Radu Iovita, Johannes Pfleging, Klaus G. Nickel, Ludovic Righetti, and Claudio Tennie. 2019. Birch Tar Production Does Not Prove Neanderthal Behavioral Complexity. *PNAS* 116(36):17707–17711. https://doi.org/10.1073/pnas.1911137116.

Schmidt, Sophie C., and Ben Marwick. 2020. Tool-Driven Revolutions in Archaeological Science. *Journal of Computer Applications in Archaeology* 3(1):1832.

Schneider, Tsim D., and Katherine Hayes. 2020. Epistemic Colonialism: Is It Possible to Decolonize Archaeology? *American Indian Quarterly* 44(2):127–148. https://doi.org/10.5250/amerindiquar.44.2.0127.

Shamay-Tsoory, Simone G., and Avi Mendelsohn. 2019. Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on Psychological Science* 14(5):841–59. https://doi.org/10.1177/1745691619856350.

Simon, Dylan A., Andrew S. Gordon, Lisa Steiger, and Rick O. Gilmore. 2015. Databrary: Enabling Sharing and Reuse of Research Video. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 279–280. Association for Computing Machinery, New York. https://doi.org/10.1145/2756406.2756951.

Skeates, Robin, and Jo Day (editors). 2019. *The Routledge Handbook of Sensory Archaeology*. Routledge, London. https://doi.org/10.4324/9781315560175.

Sonkusare, Saurabh, Michael Breakspear, and Christine Guo. 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.

Soressi, Marie, and Jean-Michel Geneste. 2011. The History and Efficacy of the Chaîne Opératoire Approach to Lithic Analysis: Studying Techniques to Reveal Past Societies in an Evolutionary Perspective. *PaleoAnthropology* 2011:334–350.

Stengelin, Roman, Robert Hepach, and Daniel B. M. Haun. 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, Dietrich. 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, Dietrich. 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, Dietrich, Thierry Chaminade, Jan Apel, Ali Shafti, and A. Aldo Faisal. 2021. The Measurement, Evolution, and Neural Representation of Action Grammars of Human Behavior. *Scientific Reports* 11(1):13720. https://doi.org/10.1038/s41598-021-92992-5.

Subiaul, Francys, Katherine Winters, Kathryn Krumpak, and Cynthia Core. 2016. Vocal Overimitation in Preschool-Age Children. *Journal of Experimental Child Psychology* 141:145–160. https://doi.org/10.1016/j.jecp.2015.08.010.

Syed, Moin, and Kate C. McLean. 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, Lucy. 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, Francisco J., Evan Thompson, and Eleanor Rosch. 2017. *The Embodied Mind: Cognitive Science and Human Experience*. Rev. ed. MIT Press, Cambridge, Massachusetts.

Weger, Ulrich W., Johannes Wagemann, and Christian Tewes. 2019. Editorial: The Challenges and Opportunities of Introspection in Psychology: Theory and Method. *Frontiers in Psychology* 10:2196. https://doi.org/10.3389/fpsyg.2019.02196.

Whittaker, John C. 1994. *Flintknapping: Making and Understanding Stone Tools*. University of Texas Press, Austin.

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

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3(1). https://doi.org/10.1038/sdata.2016.18.

Yarkoni, Tal. 2022. The Generalizability Crisis. *Behavioral and Brain Sciences* 45:e1. https://doi.org/10.1017/S0140525X20001685.