interpreting lithic raw material variability in Middle Palaeolithic contexts: a modeling approach with applications to the Bau de l'Aubesier (Southeastern France)

Pop, C.M.

Chapter Three

Re-examining basic assumptions through a systematic re-evaluation of Brantingham's neutral model of raw material procurement

Simulating Lithic Raw Material Variability in Archaeological Contexts: A Re-evaluation and Revision of Brantingham's Neutral Model

Cornel M. Pop

DOI: 10.1007/s10816-015-9262-y

Simulating lithic raw material variability in archaeological contexts:

A re-evaluation and revision of Brantingham's neutral model

Cornel M. Pop

Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology,

Deutscher Platz 6, D-04103, Leipzig, Germany. Phone: +49 (0) 341 3550 756.

E-mail: cornel_pop@eva.mpg.de

Abstract

This paper presents a systematic re-evaluation of Brantingham's (2003) neutral model of raw material procurement. I demonstrate that, in its original form, the model is ill-suited to the identification of archaeologically visible patterns, as it can only simulate processes governing the composition of toolkits and these differ substantially from those influencing the composition of discard records. I discuss and implement a series of modifications, and provide a detailed analysis of discard records produced under revised model definitions. On this basis I argue that qualitative similarities in patterns generated by the neutral model and those evidenced in archaeological contexts cannot be used to prove, or disprove, the adaptive or functional significance of raw material variability (cf. Brantingham 2003). However, I show that the revised model *can* be used to detect deviations from neutral expectations quantitatively and within well-defined error ranges.

I outline a new set of predictions for what archaeological variability should look like under the simplest procurement, transport, and discard behaviours, and argue that deviations from each of these may be traceable to specific behavioural domains (e.g. biased mobility, raw material selectivity). I also demonstrate that: a) archaeological sites or assemblages do not offer an adequate proxy for the average composition of ancient forager toolkits; b) assemblage richness is, by itself, a very poor predictor of occupational histories; and c) that the common practice of calculating expected frequencies from distances to sources is flawed, regardless of how such distances are measured.

Keywords

Lithic sourcing; raw material procurement; neutral model; agent based modeling; archaeological theory

1. Introduction

This paper presents a systematic re-evaluation of Brantingham's (2003) neutral model of lithic raw material procurement. The model, introduced as a radical alternative to traditional approaches, aims to investigate the behavioural significance of archaeological raw material variability on the basis of a minimal and well-defined set of assumptions. In essence, it is an agent-based model which simulates variability under entirely neutral (i.e. random or constant) acquisition, discard, and mobility behaviours. Surprisingly, the patterns generated by the model are qualitatively similar to those evidenced in many archaeological contexts. These similarities suggest, according to Brantingham, that the degree of representation of different raw materials in archaeological assemblages may be of "no functional or adaptive significance whatsoever" (Brantingham 2003: 490).

This argument is currently difficult to accept for at least two reasons. First, neutral explanations (i.e. that variability is the result of stochastic processes) run counter to established theory and, based on the current state of knowledge, they have a low *a priori* probability of being true. Consequently, accepting such explanations on the basis of *qualitative* similarities may be ill-advised. Secondly, Brantingham evaluates patterns evidenced in toolkit states rather than discard records, and it is unclear to what degree these are relevant to understanding archaeological variability in the first place (cf. Brantingham 2003: 493). Nevertheless, the model's ability to replicate archaeologically observable patterns cannot be ignored, and it is likely that Brantingham's work has much more to contribute to archaeological theory than a simple cautionary tale.

Despite being published over a decade ago, however, and despite its continued conceptual or applied use (e.g. Brantingham 2006; Clarkson 2007: 17; Braun et al. 2010; Perreault and Brantingham 2011; Clarkson and Bellas 2014; Lengyel 2015) and profound implications, the model itself as well as Brantingham's analyses and conclusions have yet to be systematically reassessed (cf., e.g. Oestmo et al. 2014; Duke and Steele 2010; Moutsiou 2011: 57). Thus, the goals pursued here are: a) to evaluate the representativeness of Brantingham's results, b) to ascertain the degree to which toolkit states are adequate proxies for discard records, c) to evaluate the model's ability to generate discard records which resemble the archaeological record in a meaningful way, d) to provide a detailed investigation of such discard records, and e) to discuss some of the wider implications of the neutral model.

I begin by providing a brief theoretical background in order to contextualize Brantingham's work and highlight its importance for archaeological research. Next, I assess a fully independent replication of the original model by means of 500 simulations, greatly exceeding the scope of the original study, which considered only a handful of simulations (Brantingham did not report their exact number). Through this assessment I identify a series of potential problems with the original definitions of the model, and show that the processes governing the composition of toolkits are markedly different from those governing the composition of the discard records. Moreover, I show that the original model is incapable of producing discard records that resemble the archaeological record in a meaningful way.

I implement a minimal set of revisions required to overcome the main limitations of the original model, and provide a thorough analysis of the patterns evidenced in the discard records produced by three simulations, each lasting 250 million time-steps and run with different parameters. I re-examine Brantingham's original predictions in light of my results and show that qualitative similarities between model-generated and archaeologically observable patterns cannot be used to accept, or reject, the adaptive or functional significance of raw material variability (cf. Brantingham 2003). However, I also show that the revised model can, theoretically at least, detect deviations from neutral expectations with a relatively high degree of precision and known, well-defined error ranges.

My results do not support some of Brantingham's original predictions, but they do provide a new set of expectations for what variability should look like under the simplest procurement, transport, and discard behaviours. I argue that predictions made on the basis of these expectations may be rejected on an individual basis, and that deviations may be traceable to variables influencing particular behavioural domains (e.g. resource selectivity, mobility). Ultimately, the revised model presented here is also, by necessity, a model for the formation of the archaeological record, and I show that it can be used to investigate issues of fundamental importance for archaeological research, including the relative importance of lithics in past adaptations and the meaning of archaeological sites. I therefore conclude my discussion with a brief examination of the wider theoretical implications of the model.

2. Theoretical background

2.1. Lithic procurement studies in an evolutionary context

Stone is the only resource exploited throughout human history which is both well-represented in most archaeological contexts and whose geographic provenance can be ascertained with some confidence. The representation of different raw materials can therefore be evaluated in terms of the location and characteristics of sources in a given area, allowing for useful information on human/landscape interactions to be recovered at most archaeological sites.

In principle, this information allows for insights into past landscape utilization patterns (e.g. Binford 1979 and Gould and Saggers 1985; Gould 1978; Féblot-Augustins 1993, 1999; Fernandes et al. 2008; Brantingham 2006; Arakawa 2006), social behaviours (e.g. Pearce and Moutsiou 2014; Moutsiou 2011; Gould and Saggers 1985), and even the cognitive abilities of extinct hominins (e.g. Moutsiou 2011; Braun et al. 2009; Braun et al. 2008; Spinapolice 2012 and references therein). Additionally, it allows for more refined interpretations of other aspects of stone tool variability, since it is generally acknowledged that technological choices, the final shape of artifacts, and the use-lives of different tools can be affected, to varying degrees, by the intrinsic properties (e.g. hardness, nodule shape) and the availability of different raw materials (Dibble 1991; Kuhn 1991; Andrefsky 1994, 2009 and references therein; Driscoll 2010).

2.2. Challenges of modeling past procurement behaviours

Procurement studies can play a key role in shaping our understanding of the evolution of the human condition, but only if adequate theoretical frameworks exist for inferring dynamic behaviours from what is in many ways a static record. Building theoretical models of raw material procurement is a difficult task, however, and choosing between alternative explanations is often a challenge. The reasons for this are numerous, but the following deserve some brief consideration:

a. Raw material procurement may have been embedded in other activities (e.g. Binford 1979), potentially unrelated to basic subsistence[1], and in many cases it may be impossible to determine whether, or to what degree, that was the case (Gould and Saggers 1985: 123; cf. Myers 1986: 66-72). Thus raw material variability may have been governed, to varying degrees, by factors unrelated to lithic procurement proper.

b. The spatial range of lithic exploitation need not necessarily correspond to the range over which other resources were procured. This possibility renders inferences about mobility or foraging radii difficult to draw (Cole 2002; cf. Myers 1986: 60-65).

c. We lack adequate ethnoarchaeological information on how lithic resources were managed by foragers who still depended on them in modern times[2]. It is therefore difficult to develop well-grounded theories of how raw materials may have been procured and used in the past, even when investigating traces left by anatomically modern humans.

d. Past choices are difficult to evaluate because both the cost and the benefits of procuring specific raw materials are hard to determine. First, "cost and distance to source are not equivalent" (Kuhn 1991: 78), not least because cost depends on the specifics of spatial practices (i.e. land use, mobility) as well as conceptualizations and perceptions of space (*sensu* Lefebvre 2007). Second, the quality (i.e. flakeability, durability) or benefits afforded by different materials are both contextually-dependent (e.g. what use a tool was put to) and notoriously difficult to quantify (e.g. Woods 2011: 84-99; Pop 2013). Indeed, quality may have been, and likely was, considered from a simple binary perspective: either good enough, or not, for whatever purpose was envisioned at the time.

In practice these difficulties lead to a series of fundamental but hard to evaluate assumptions being built into virtually all current models of raw material procurement. These assumptions can manifest in the choices of variables to be used (e.g. distance to source), in the differentiation between physical and socio-cultural factors (e.g. definitions of 'cost' and 'quality'), in notions of what archaeological sites represent (e.g. periods of continued habitation), and even in our perceptions of the relative importance of lithics in hominin adaptation. The fact that potentially unfounded assumptions are built into our models does not imply, of course, that these models are of no value. On the contrary, such models have significantly enriched our understanding of the past (e.g. Adams and Blades 2009; Yamada and Ono 2014). The inclusion of these assumptions does imply, however, that the reliability of inferences derived

[1] Binford (1979: 259) argued that lithic raw material procurement was embedded in basic subsistence activities, but there is no reason to suspect that it may not also have been embedded in non-subsistence activities which may not be visible archaeologically (e.g. Gould and Saggers 1985: 122).
[2] While there are a series of ethnoarchaeological accounts of lithic procurement (e.g. Binford 1979; Gould and Saggers 1985; see also the review by McCall 2012), the studied populations had access to modern means of transport (e.g. Gould and Saggers 1985: 120), and even in the case of Binford's study of the Nunamiut, which provided the basis for his "embedded procurement" argument, it failed to consider variables that are critical in understanding procurement behaviours (see Cole 2002: 54).

solely on the basis of the provenance of stone is questionable and, moreover, that we are limiting our inquiry to variables that may, ultimately, not be the most significant in explaining past behaviours.

2.3. Brantingham's (2003) neutral model in context

Brantingham's work is valuable in this context because it provides an alternative starting point, constructed on the basis of minimal and explicitly defined assumptions. In essence, it aims to provide insight into what variability should look like, and more importantly perhaps, why, if only the effects of random processes and constant parameters are considered. Briefly, Brantingham's model aims to be:

a. *A neutral model*: Neutral models are similar, but not identical to, null models, of which sometimes they are considered a special case (Gotelli and McGill 2006; cf. Brantingham 2003: 490-491). They are defined by Pearson and Gardner (1997: 215) as "a minimum set of rules required to generate pattern in the absence of a particular process (or set of processes) being studied." In essence, they aim to evaluate the effects of "noise", or those processes which are not relevant to the questions being asked, so as to more reliably detect "signals". In the case of Brantingham's (2003) work these questions pertain to adaptive processes and, consequently, his model seeks to evaluate the effects on non-adaptive behaviours.

b. *An agent-based model:* Agent-based models aim to computationally simulate complex systems on the basis of the local interactions of autonomous, decision-making agents (e.g. Bonabeau 2002; Janssen 2012). They are useful because: a) the emergent, macro-level patterns generated by these low-level interactions are often difficult or impossible to predict through traditional analytical approaches, and b) by modifying the set of behavioural rules each agent is given it becomes possible to study the effects of isolated variables and specific parameters on the behaviour of the overall system.

In following an agent-based approach Brantingham (2003) encodes neutral processes in a set of behavioural rules. Specifically, he simulates a single agent which: a) has a fixed capacity to carry raw materials; b) randomly moves to an adjacent cell or stays in place at every simulation time-step; c) collects as many raw materials as possible upon chance encounters with sources; and d) if possible (i.e. if its toolkit is not empty), discards one raw material unit at random after every move. The landscape with which the agent interacts contains randomly distributed sources of identical quality (however defined) and infinite, uniformly distributed non-lithic resources.

There are many advantages to such agent-based neutral models. These include simplicity, rigorous formal definition, and the ability to: a) detect and understand emergent patterns, b) study the systemic effects of individual variables, and c) provide well-defined null-hypotheses against which archaeological variability can be evaluated. In principle, then, Brantingham's model provides a strong foundation for building bottom-up interpretive frameworks that offer an alternative to traditional top-down approaches.

It must be noted, however, that Brantingham's model is not, strictly speaking, a neutral model: as suggested by Brantingham himself (2003: 504), the modeled behaviours (e.g. random procurement) may have been adaptive, and consequently there is no clear distinction between what constitutes "noise" and “signal”. Beyond this, neutral models require specifying parameters which may be impossible to estimate directly (see Gotelli and McGill 2006), and as such are neither assumption-free nor particularly reliable. Indeed, fiddling with parameters can lead to a variety of patterns, including some that may resemble those observed in real contexts (Enquist et al. 2002). Finally, Brantingham’s model is only agent-based in a restricted sense, since interactions between agents, or indeed, the effects of simulating multiple agents, are not considered.

2.4. Theoretical implications of the neutral model

Brantingham (2003) argued that the neutral model is capable of replicating a number of commonly observed archaeological patterns. If correct, this has a number of important implications for archaeological theory. The most direct implication, and one which is discussed by Brantingham (2003) in detail, is that raw material variability may not be informative with regard to past adaptation. I consider this conclusion to be overly pessimistic but, if correct, it would imply that our ability to extract useful information from the representation of raw materials in archaeological contexts is severely limited. It would also imply, of course, that many current interpretations are wrong.

The implications of the model extend far beyond the issue of raw material procurement, however. For instance:

a. If the neutral model accurately simulates procurement behaviours we would have to accept the possibility that access to raw materials, and consequently access to stone tools, was far less important in the past than commonly believed[3]. This is because the requirements of random walks are incompatible with the requirements of constant access to lithic sources. Indeed, Brantingham (2003: 494) reports that even in the context of a simulation in which the average distance between raw material sources was but 4% of the maximum distance that the agent could traverse with a full toolkit, said agent spent 25% of simulation time without any access to stone tools.

b. A good correspondence between the model’s output and patterns observed in archaeological assemblages also suggests that lithic recycling played at best a very minor role (cf. Amick 2007).

c. Such good correspondence would also imply that sites were never a living space or dwelling of any kind - under the model, assemblages are simply random aggregates of items discarded over time by agents whose probability of remaining at any one location for two consecutive time-steps is only one in nine.

Taken together, and based on our current state of knowledge, these implications suggest, but do not demonstrate, that if it accurately simulates archaeological patterns the neutral model is probably wrong. Insofar as the model is taken to be an exploratory tool rather than a realistic simulation of the past, however, this may be largely inconsequential. Indeed, it would be expected that attempts to

[3] As discussed by Sillitoe and Hardy (2003), there is certainly evidence to suggest that we may be overestimating the “true place [of lithics] within the material culture of which they formed part” (Sillitoe and Hardy 2003: 555).

demonstrate the invalidity or incorrectness of the model would result in productive reassessments of some fundamental assumptions regarding the very nature of the archaeological record.

3. Methods

In order to avoid perpetuating potential errors, I replicated Brantingham's model solely on the basis of the definitions provided in the original publication (Brantingham 2003: 491-494), without examining Brantingham's code or that of alternative implementations (Janssen and Oestmo 2013). I coded the model in Python (v3.4.3) with the aid of the SciPy (v0.15.1) and NumPy (v1.9.1) libraries (van der Walt et al. 2011), relying on SciPy's implementation of the "Mersenne Twister" algorithm (Matsumoto and Nishimura 1998) for the generation of pseudo-random numbers. I tested both the Python code and the simulation outputs for correctness and compliance with the definitions of the model (see Table S2 in Electronic Supplementary Materials [ESM] 1).

I analyzed the outputs of 500 simulations of the original model, focusing on the variability and patterns evidenced across the entire simulation set (c.f. Brantingham 2003). This number was chosen arbitrarily, but it was deemed large enough to offset potential biases resulting from variations in the local density of the randomly placed sources and the expected low degree of coverage attainable by agents in individual simulations. In this regard it should be noted that none of the simulation lengths reported by Brantingham (2003) exceeded 25,000 time-steps, or 10% of the minimum that would be required to achieve complete coverage of the simulated world (a 500x500 grid, containing 250,000 cells). The low degree of coverage, coupled with the random (i.e. non-uniform) placement of sources, renders individual simulations susceptible to biases resulting from local differences in source densities.

My analyses of the replicated model indicated that the termination conditions specified by Brantingham end simulations prematurely, thereby biasing overall results and also preventing the formation of discard records suitable for detailed investigations. I therefore modified the definitions of the model without fundamentally altering it. The algorithm of the revised model is outlined in Figure 1 (cf. Brantingham 2003: 494), and the modifications are summarized below:

a. I removed the condition upon which simulations terminate after agents encounter 200 unique raw material sources. Instead, I define a new, single termination condition whereby simulations end after a pre-defined number of time-steps. This allows for the results of simulations to be interpreted in terms of patterns evidenced within a specified time-span, without having to consider the effects of the variable time required to achieve a specific degree of coverage of the simulated world (e.g. reaching a set number of sources). Thus, this single termination condition allows for easier comparison of simulations run with different mobility, discard, and toolkit size parameters.

b. I modified the behaviour of agents at the edges of the simulated world so as to allow for arbitrarily long simulations. In the revised model agents continue to have an equal probability of moving to one of the adjacent cells or staying put, but only those cells that exist within the confines of the world are considered: thus, four movement options are available to agents at the corners of the grid, six at non-corner edge cells, and nine at any other grid location.

c. I slightly modified the logical sequence of the algorithm so as to allow agents initially placed on a source to fill their toolkit in the first simulation time-step (cf. Brantingham 2003: 494).

d. I optimized the model for running long simulations (hundreds of millions of time-steps), partially by discarding information on toolkit states[4] and retaining only information on the discard record.

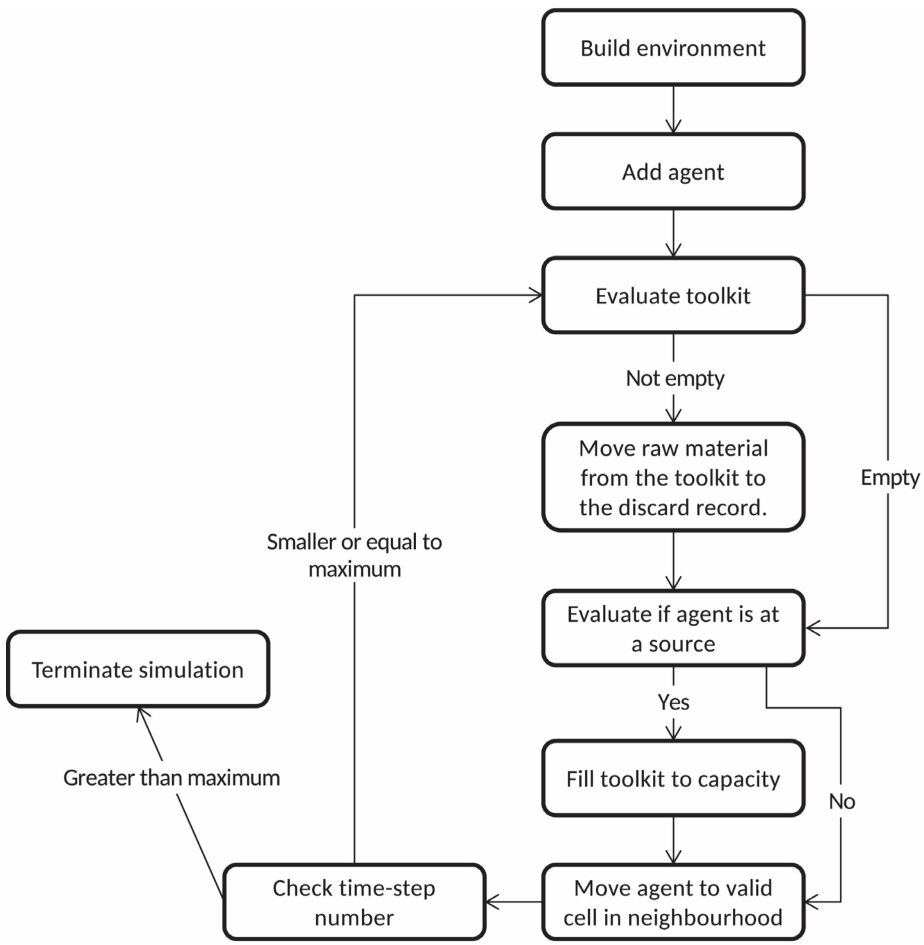

Fig. 1 Flowchart of the revised model's algorithm. Simulations only terminate when a pre-defined maximum number of time-steps has been reached

As with the replicated model, the outputs of all 3 main and 9 auxiliary simulations (250 million time-steps each – see sections 4.2 and 4.2.2.2 for an explanation of the number, and lengths, of simulations) run under revised model definitions were tested prior to analysis. The tests were very similar to those summarized in Table S2 (ESM 1), but differed in two ways: first, due to computational constraints simulations of the revised model only recorded toolkit state data for the initial 200 thousand time-steps,

[4] Information on toolkit states was still recorded for the first 200 thousand time-steps, but only for diagnostic purposes.

and consequently some tests could only be conducted on a small portion of the simulation outputs; secondly, some tests had to be modified to account for differences in model definitions (e.g. termination conditions).

All diagnostic tests and statistical analyses were conducted in R version 3.1.2 (R Core Team 2014) with the aid of RStudio (RStudio 2014) and a number of R packages, namely: nabor version 0.4.3 (Elseberg et al. 2012) for fast KD-Tree nearest neighbour searches, data.table version 1.9.4 (Dowle et al. 2014) for fast manipulation of the large simulation data sets (>4GiB per simulation under revised model definitions), raster version 2.3-33 (Hijmans and van Etten 2012) and rasterVis version 0.35 (Lamigueiro and Hijmans 2014) for the generation, manipulation, and visualization of raster data, spatstat version 1.39-1 (Baddeley and Turner 2005) and sp version 1.0-17 (Bivand et al. 2013) for simplified analysis and manipulation of spatial data, and testit version 0.3 (Xie 2013) for running diagnostic tests on simulation outputs.

The R code for all analyses and plots, as well as the Python code for the revised and replicated models are available under the terms of the GNU General Public License version 2 or later in the Electronic Supplementary Materials (ESM) 2 and at: http://alyanne.net/software/SABM/.

4. Results

4.1. A re-examination of Brantingham's neutral model

4.1.1. <u>General observations on replication results</u>

The general toolkit patterns discussed by Brantingham (2003) were replicated successfully (see Supplementary Electronic Materials [ESM 1] for a full account). However, while all values reported by Brantingham fell within the range of variability evidenced in the outputs of the 500 simulations that I analyzed, they were seldom truly representative. Most of the observed discrepancies are likely explainable by the fact that the lengths and outputs of the simulations are highly variable. This variability, which entails a low probability of observing representative values across all investigated relationships in the context of a single or small number of simulations (Brantingham analyzed only a handful of simulations, but the exact number was not reported), is mainly the result of the arbitrary termination conditions, which cause simulations to end prematurely (see ESM 1 for more details). Some minor discrepancies can also be attributed to problematic aspects of Brantingham's analysis, however. Specifically, linear distances, which are incompatible with the definitions of Brantingham's model, are employed in derivations of foraging radii and maximum expected toolkit richness values; as a consequence, both derivations are incorrect (see ESM 1 for details).

The results of my replication call into question some of Brantingham's conclusions. In particular, I was unable to confirm Brantingham's postulated relationship between modal and maximal transport distances, or that between median/modal transport distances and the sizes of the foraging radii (Brantingham 2003: 498-500; see ESM 1 for more details). This renders Brantingham's analysis of archaeological parallels debatable, because the reported ratios of maximal to modal transfer distances

(2 for the Aquitaine Basin, 4.6 to 6 for the Middle Palaeolithic of Central Europe – see Brantingham 2003: 503) are much lower than the ratio observed in my simulations (approximately 20) [5].

My analyses also confirmed several of Brantingham's conclusions, however. These include: a) the assertion that the neutral model predicts a strong correlation between maximum toolkit richness states and toolkit sizes (Brantingham 2003: 498), b) the observation that modal and median transport distances observed in toolkits are minimally affected by resource densities (Brantingham 2003: 499), and c) the observation that long distance raw material transfers can be explained by very simple behaviours (Brantingham 2003: 503).

4.1.2. Toolkit and discard record patterns

The combined output of the 500 simulations revealed important differences between the patterns evidenced in the simulated toolkits and those observable in the discard records, or the records of the final resting locations (grid cells) of all items removed from the toolkit. These discrepancies, which are discussed in detail below, call into question Brantingham's assertion that "[i]ndividual archaeological assemblages may be treated as repeated random samples of different sizes from the mobile toolkit" (Brantingham 2003: 493)[6], at least insofar as such samples are not spatially constrained[7], and the archive of discard events can be considered a more pertinent analogue for the archaeological record than ephemeral toolkit states. This, in turn, calls into question the relevance of simulated toolkit patterns for understanding archaeological variability.

The most dramatic difference was observed in the relationship between sample size (i.e. the total number of items per cell or toolkit state) and raw material richness (i.e. the number of unique raw materials represented). As shown in Figure 2a, the discard records evidence a clear increase in richness states with increasing sample sizes, with observations most frequently displaying a 1:1 or similar ratio. In the toolkit data set (Figure 2b) a small increase in richness states with increasing sample size can also be observed, but the nature of the overall relationship is much less clear. Importantly, the toolkit data

[5] Note that Brantingham's own reported modal and maximum transport distance values are also inconsistent with his predicted ratio of 3 to 4. Indeed, Brantingham (2003: 499) first argues that "the neutral model predicts that maximum raw material transport distances should be three to four times the *foraging radius d*" (italics mine), but in his discussion of archaeological parallels he states that "[m]aximum transport distances are expected to be three to four times the distance represented by the *internal mode*" (Brantingham 2003: 501; italics mine). It is unclear why Brantingham equates foraging radii with internal modes (Brantingham 2003: 503), given that the foraging radius value he uses is 10 (e.g. Brantingham 2003: 499) and the modal value is 5 (Brantingham 2003: 498). Using the actual modal value he provides, the maximal to modal distance ratio is 8.6, not the 3-4 value used in his analysis of archaeological parallels.

[6] Note that the model provides only one and very weak mechanism for allowing such "random samples of different sizes (Brantingham 2003: 493) to be drawn from the toolkit, namely the 1/9 probability that the agent will remain at the same location for two consecutive time-steps, discarding two items from the same toolkit at the same location. While arbitrarily obtaining toolkit samples larger than one by simply querying it for multiple items is technically possible, doing so would violate the definitions of the model (i.e. discard rates) and would provide meaningless results.

[7] It should be noted that Brantingham discusses patterns evidenced in toolkits states sampled over time, not in space (i.e. at a specific grid coordinate). This is probably because, as shown in the text, the definitions of the model prevent such spatial sampling.

points tend to cluster at high sample size values across all richness states, while in the discard record there appears to be a limit on the *minimum* richness states observable with high item counts; Indeed, no cells containing 24 items or more were found to have less than 3 unique raw material types represented (Figure 2). Moreover, it should be noted that the sample sizes for individual cells are low, never exceeding one third of the size of the mobile toolkit despite the fact that individual simulations of over fifty thousand time-steps were observed, and the cumulative total for all simulations exceeded 8.1 million. Given these differences, it is clear that random samples from the toolkit data set provide only a distorted view of the patterns evidenced in the discard record.

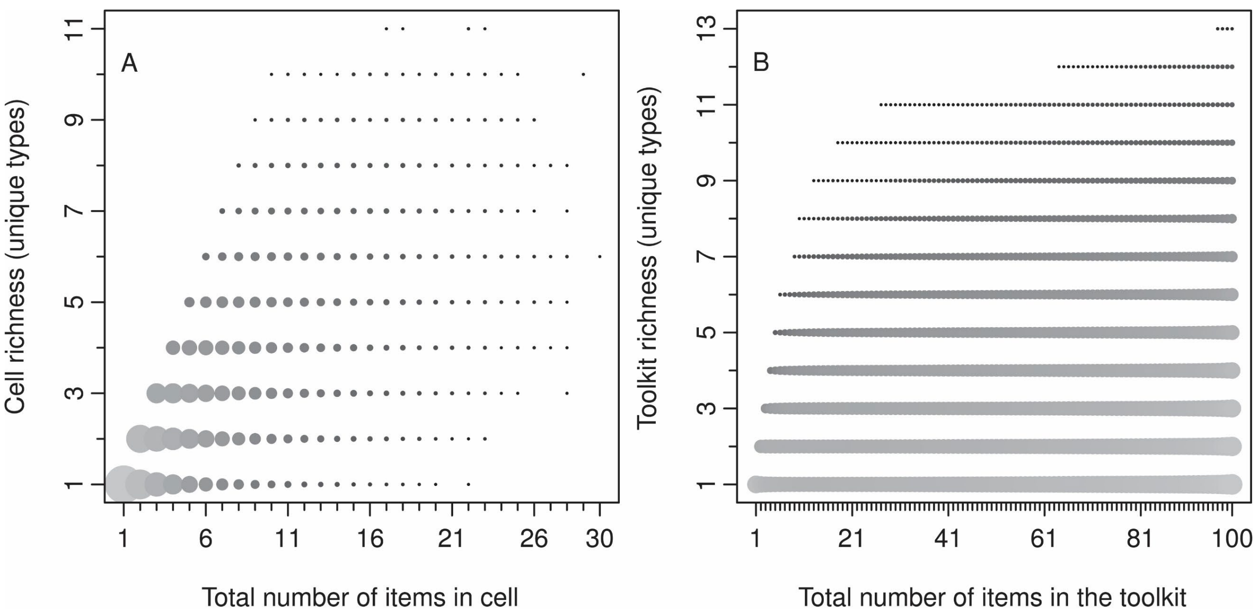

Fig. 2 Relationship between richness state (number of unique raw material types) and sample size in a) the discard record (n=1.92 million), and b) toolkit states (n=8.1 million). Filled circles denote individual variable combinations, while their size and grayscale values represent the number of individual observations. Circle diameters grow with the fourth root of sample size ($n^{1/4}$), so as to permit visualization despite the extreme differences in the frequencies of observation (1 to 700586); colours, on the other hand, are given as the log of sample size transformed to fit in the range of 0 (black) and 0.75 (light gray)

Notable albeit less pronounced differences in transport distance patterns can also be observed. As shown in Figure 3, although the modal transport distance remains the same across both data sets (3 grid cells), the probability of observing it is approximately 15 percent lower in the discard record; at the same time, the probability of observing long transport distances is higher than that recorded in the toolkits despite the considerably lower sample size (5.3 million versus 333.7 million).

These differences may seem surprising given that both data sets encode different aspects of the same information, but they are explainable by two simple facts. First, toolkits are mobile and encode primarily changes over time, while discard locations are fixed and encode mainly changes over space. Consider, for example, a situation in which four sources are placed at an equal distance of five grid cells from a given location *x*, but in different directions (e.g. north, south, east, and west). Because the probabilities of observing transport distances of 5 grid cells or more are high (5 grid cells corresponds to the median transport distance - see ESM 1), we should expect all four sources to be well represented at location *x*, provided that simulations are long enough to allow the agent reasonable coverage of the simulated world. Conversely, the probability of observing all four sources being represented in a single snapshot of the toolkit is low, because the agent would have to continuously carry raw materials over the minimum distance of 20 grid cells that is required to visit all four sources[8]. Thus, all else being equal (e.g. equal sample size), we would expect the raw material richness of the discard record to be higher than that observed in the toolkits.

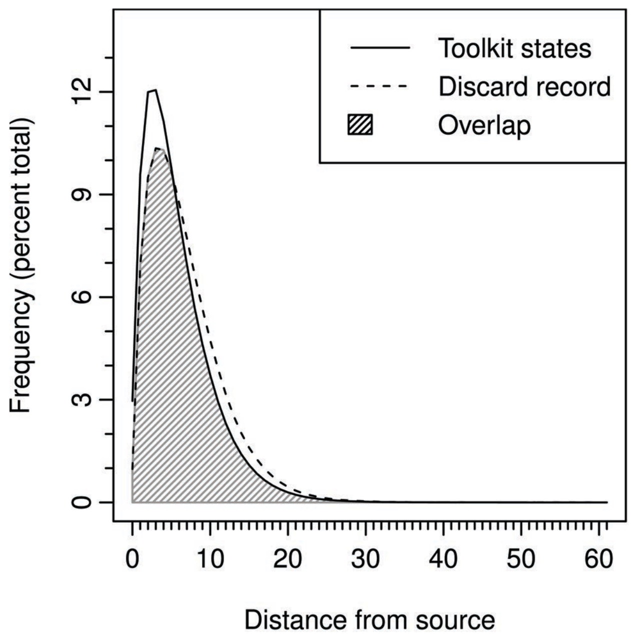

Fig. 3 Comparison of transport distance frequency distributions evidenced in the aggregated toolkit and discard record data sets. Observations on all transported items across the 500 simulations (8.1 million time-steps) are included; these total 333.7 and 5.3 million in the case of toolkit states and discard records respectively. The large differences in sample sizes are due to the fact that only one item may enter the discard record per time-step, while toolkits can and often do contain multiple items at any given point in time. Frequencies are given as percentages of the total number of observations for each data set

A second, compounding factor is that there are no upper limits to the number of items that may enter the discard record at a given location, except those defined by the simulation termination conditions. Input to the toolkit, however, is limited at any given point to its overall carrying capacity of 100 raw material units. Consequently, richness states greater than 100 raw material types could never be observed in the toolkits, even within the context of simulations run with an extreme density parameter of one; In the discard records, on the other hand, such richness states would be expected to be common, because a source density of one implies that 100 distinct raw material sources are always found within a radius of 5 grid cells from any given location, or a distance that corresponds to the

[8] The distance an agent may cover in a given number of time-steps is equal to the number of time-steps, regardless of direction, because diagonal movements carry no penalty; in other words, under the mobility rules stipulated by Brantingham, the hypotenuse of a right triangle is equal in length to the other two sides. Thus, the distance between two sources located at x = ±5 and y = ±5 from each other is 5 grid cells.

median transport distance observed across my simulations (see ESM 1). Toolkits can therefore offer only a biased view of maximum attainable richness states, and indeed richness states in general.

The fill rates of toolkits are also variable, and generally high (1-100 raw material units, depending on how full the toolkit is when a source is encountered), while those of the discard record are constant and low: only one raw material unit may be discarded in any given time-step. These variable fill rates bias the transport distance patterns recorded in the toolkits towards lower values, because up to 100 observations of transport distances of zero may be recorded in a single time-step during source encounters, while only one such observation may be recorded in the discard data. In other words, the relative frequencies of specific transport distance observations at or near to raw material sources is higher in the toolkits than it is in discard records, and this translates to different probability frequency distributions.

In short, the processes governing the composition of toolkits differ from those governing the composition of the discard record, and consequently the former are poor proxies for the latter. The opposite is also true, of course: discard records, and by implication archaeological sites or assemblages, do not and cannot offer, as it is all too often assumed, an adequate proxy for the average composition of ancient forager toolkits. Critically, this observation remains true even if factors such as post-depositional processes and time-averaging could be assumed to have played no role in shaping archaeologically observable patterns.

4.1.3. Limitations of Brantingham's neutral model

As noted in the previous section (see Figure 2a), the maximum number of items discarded at a single location across all 500 simulations is very low (30 raw material items), even relative to the carrying capacity of the simulated agents (100 raw material units). This is primarily a consequence of the arbitrary end conditions stipulated by Brantingham (i.e. encounters with 200 unique sources or with the edges of the world), which prevent the simulated agent from revisiting specific locations a large number of times (a maximum of 40 cell visits were recorded in individual simulations). However, the low rates of discard (one item per time-step), as well as the random mobility of the agents are also contributing factors.

As shown in Figure 4, the maximum number of items discarded at a single location increases non-linearly with simulation length. To assess the exact nature of this relationship I used a Generalized Linear Model (GLM) with Poisson error structure and log link function (McCullagh and Nelder 1989). I set the predictor variable to the minimum number of time-steps required to produce a given cell size in the context of individual simulations[9], log transformed to ensure a balanced distribution of observations. The response variable, on the other hand, represents the maximum cell sizes evidenced at different points[10] along the simulations. Although leverage values indicated the presence of some influential

[9] For the purposes of the GLM simulation time-steps were counted from the first discard event onward; thus, the first discard event represents the first simulation time-step.
[10] Cell size was increased by one whenever a given time-step resulted in a higher maximum cell size than previously observed.

cases, removal of these did not affect the overall results (a difference of one percent or less in coefficient estimations, and no appreciable difference in *p* values); moreover, all other regression diagnostics were found to be within acceptable ranges, indicating that the model fits the data well[11].

The resulting regression formula indicates that maximum cell richness grows exponentially at a rate of 0.546 (SE = 0.0036, $p < 2*10^{-16}$) times the common logarithm of simulation length (intercept = 0.697 [SE = 0.0122, $p < 2*10^{-16}$]). This indicates, in turn, that based solely on the observed relationship between these variables simulations lasting approximately 14.5 million time-steps (i.e. $10^{7.16}$) or more would be required to generate cells (i.e. sites) which match or exceed the size of toolkits. More importantly, however, the formula indicates that approximately $10^{19.6}$, or ca. 50 quintillion (short-scale) time-steps would be required to generate cells comparable in size to sites such as Pech de l'Aze IV, where approximately 92000 lithic artifacts had been recovered by the end of the twentieth century (McPherron and Dibble 1999). Such simulation lengths are, unfortunately, virtually impossible to attain under the original definitions of the neutral model (the maximum length observed across all the 500 simulations was 50866 time-steps).

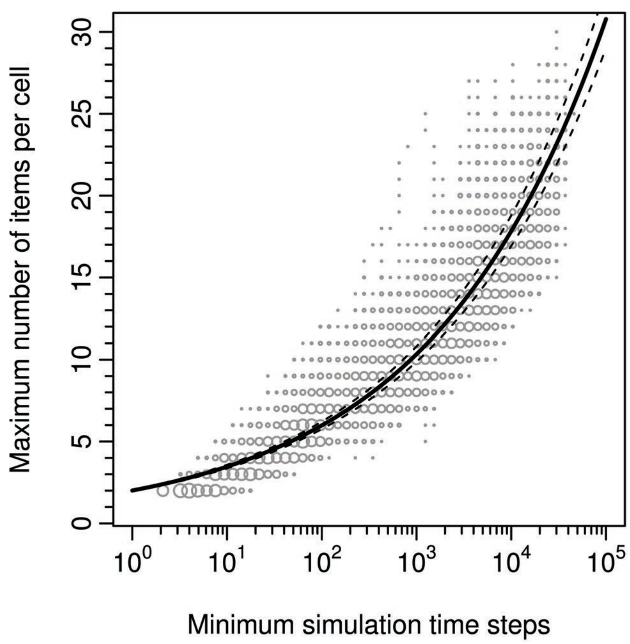

Fig. 4 Relationship between maximum cell sizes (i.e. number of items discarded at a single location) and the minimum number of simulation time-steps required to attain them in the context of individual simulations. A total of 7287 observations from 500 simulations are included. For display purposes observations are grouped into 50 bins along the *x* axis, and are represented by circles, whose areas indicate relative differences in the number of observations per bin at each *y* value. The regression line, defined by the equation $y = e^{0.697+0.546*log10(x)}$, is indicated in solid black, while the 95% confidence interval is identified by dashed lines. Note that the maximum value along the *y* axis exceeds the maximum recorded simulation length of 50866, so as to show the predicted minimum number of time-steps required to achieve the maximum observed cell size of 30

[11] The Python code for the generation of the raw data as well as the R code for all analyses are included in the Electronic Supplementary Materials (ESM2).

It could be argued that sampling toolkits for multiple items at specific spatial locations (i.e. grid cells) would yield larger sample sizes than those evidenced in the discard record. While this is true, such sample sizes would still fail to approach the size of many archaeological sites[12], and would in any case violate the definitions of the model (one and only one item may be removed from the toolkit at any one time-step). It could also be argued that simulating multiple foragers, either sequentially or in parallel, would result in the formation of much larger assemblages. This is also true. However, doing so would result in spatially biased outputs because the probabilities of termination vary across the simulated world: an agent cast near the edges of the world has a higher chance of being terminated quickly than one cast at the center of the grid, and consequently cells in the center would be visited more frequently than cells near the edges.

Given the preceding discussions it is clear that the original definitions of the neutral model render it ill-suited for its stated objective, namely the study of archaeologically visible patterns generated under the simplest procurement, transport, and discard behaviours. Nevertheless, all issues discussed above can ultimately be traced to a single design factor[13], namely the arbitrary conditions under which simulations terminate; consequently, removing these and forcing agents to remain within the confines of the simulated world would address all identified limitations. As shown, however, revising the original definitions is not enough – a change of focus from toolkit states to discard records is also necessary, because the patterns evidenced in the two data sets are not identical.

4.2. Discard records under a revised neutral model

I implemented the revisions suggested above and investigated the discard records produced under revised model definitions (see the Methods section) through three main simulations. Each of these simulations, henceforth referred to as A, B, and C, ran for a total of 250 million time-steps and allowed for an average of 1000 agent visits per cell. For simulation A Brantingham's default parameters were used (see Table S1 in ESM 1), while simulation B was run with a modified resource density of 0.002 (i.e. 500 sources). For simulation C I used the same resource density as in simulation B, but modified the carrying capacity of the toolkit to 200 units, therefore altering the toolkit size to discard rate (T/D) ratio; all other simulation parameters were kept constant across the three simulations.

[12] Multiplying the maximum cell site observed across all 500 simulations (30) by the maximum size of the toolkit (100) would yield but 3000 items.

[13] Two other aspects of the model may also prove problematic. First, random mobility cannot guarantee access to lithic sources (or indeed sites), and as argued by Oestmo et al. (2014) and also shown by my analyses (see ESM 1), this brings into question the survivability of foragers. However, the model makes no provisions for lithic recycling, which would allow foragers to recover previously discarded items, and it is in any case unclear whether access to lithic materials was as critical in the past as commonly assumed. While future revisions of the model could, and should, include provisions for non-random or semi-random mobility, whereby agents are allowed to gravitate around resources or sites, random mobility can nevertheless be accepted as a baseline. The second problematic aspect pertains to the number of modeled foragers: clearly, a single forager was not responsible for the formation of the archaeological record, and inter-agent communication likely played a major role. However, preliminary analyses revealed no differences between models run with multiple or single agents, and allowing for inter-agent communication would dramatically increase the complexity of the model. Therefore, as with random mobility, the patterns generated by a single agent provide an adequate baseline for interpreting the record and evaluating the effects of new variables that may be included in the model at a later stage.

My analyses focus on two different but complementary aspects of the discard records, namely raw material transfer and source representation patterns. I discuss these both at a macro level (i.e. the average patterns that would be observable in the aggregated data from multiple randomly selected sites) and at the level of individual cells (i.e. the patterns observable at specific locations). The patterns evidenced in the outputs of individual simulations are compared in order to determine the influence of the two non-constant model parameters (resource densities and toolkit sizes) at both levels of analysis. The results of this investigation are described in detail below.

4.2.1. General characteristics

The discard records generated by the three simulations (Figure 5a1,b1,c1) attained comparable maximum cell (i.e. site) sizes, exceeding the maximum observed in simulations run under original model definitions by a wide margin (918 to 1234 versus 30 items per cell). However, they varied greatly in terms of overall richness and in terms of the number of unique raw material types represented in individual cells (Table 1). The records share three general characteristics:

a. In all three discard records the quantity of items found in individual cells decreases rapidly with increasing distance from the nearest raw material source(s). This indicates that, regardless of what parameters (e.g. toolkit sizes, source densities) are used, the modeled behaviours are only capable of generating large "sites" at or very near to raw material sources. Importantly, this spatial pattern is not explicable in terms of biases in agent mobility or discard behaviours[14].

b. Some regions of the simulated world show a higher overall density of discarded items than others, notwithstanding any similarities in the distribution of sources. This occurs due to random variations in the number of agent visits per cell (see Figure 5a2,b2,c2), which in simulation A varied between 334 and 1378 (mean = 1000, SD = 95.6, matching theoretical expectations – i.e. the number of simulation time-steps divided by the number of grid cells, or the average predicted number of cell visits), and indicates that the effects of random walks must be carefully considered when investigating the compositional variability of individual cells.

c. Across all three simulations the contents of individual cells grew linearly with simulation length, but at a highly variable rate. In simulation A, for example, the contents of the largest cell grew at a rate of approximately one item every 200 thousand time-steps, while the contents of the smallest cell grew at the much slower pace of only one item per 7.2 million time-steps; in both cases the correlation between the variables was very strong (r^2 = .9971, n = 1234, $p < 2*10^{-16}$ and r^2 = .925, n= 38, $p < 2*10^{-16}$ respectively).

[14] The frequency of moves in each of the 9 possible directions was consistent with those expected under conditions of random draws from uniform distributions (Simulation A: χ^2=2.8047, df=8, p=0.946; Simulation B: χ^2=6.9466, df=8, p=0.542; Simulation C: χ^2=5.1084, df=8, p=0.746); the same was true with regards to the selection of items for discard: the distribution of *p* values for the observed versus expected frequencies of selected discard indices at different toolkit sizes was uniform across all three simulations, with Simulations A, B, and C yielding significant tests in 4.04, 5.05, and 3.02 percent of cases respectively at an alpha level of 0.05.

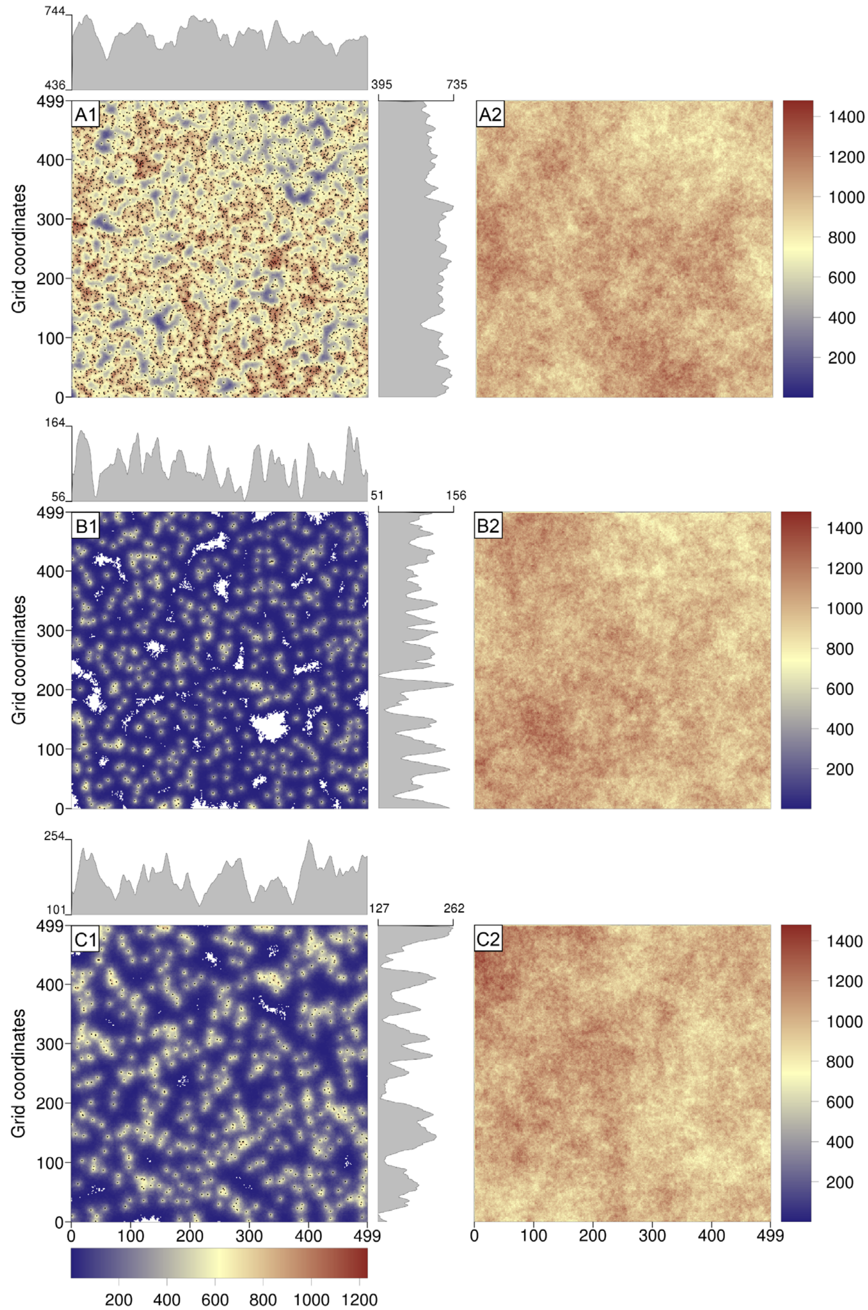

Fig. 5 Discard records (*a1, b1, c1*) generated by simulations A, B, and C, indicating the total number of items found at each grid location after 250 million time-steps. Raw material sources are indicated by black dots, while the gray margin plots show the mean number of discarded items per cell along the *x* and *y* coordinate axes. The colour scheme and legend for all discard record plots are identical. Total agent visits to individual grid cells are shown in subplots *a2*, *b2*, and *c2* for simulations A, B, and C respectively. The colour scheme and legends are identical across all three grid coverage plots

This is an important observation because the deterministic component underlying the observed relationship between cell sizes and simulation lengths reflects the spatial relationship between cells and sources in their neighbouring area, mediated by the transport capacity of the agent (see below). Thus, discarded materials can accumulate in cells (i.e. sites) at vastly different rates even if visited the same number of times by an agent(s) in the context of simulations run with identical population and behavioural settings (i.e. one agent, following constant behavioural rules). This strongly suggests that assemblage richness is, by itself, a very poor predictor of demographic parameters; consequently, unless the influence of the spatial distribution of sources is taken into account, very little (if anything) can be inferred in terms of differences in population history from differences in the size of assemblages or sites, even if it can safely be assumed that technologies, curation strategies, and sedimentation rates were constant.

Table 1 Summaries of simulation parameters and generated discard records. Model parameters not shown here are identical to the defaults used by Brantingham (2003) and specified in Table S1 of the Electronic Supplementary Materials (ESM 1)

Variable	Simulation A	Simulation B	Simulation C
Number of sources	5000	500	500
Toolkit size	100	100	200
Total number of discarded items	165,153,161	25,476,408	44,545,914
Discarded items relative to theoretical maximum (250 million)	66.06%	10.19%	17.82%
Cells with discarded items	100%	95.14%	99.52%
Minimum number of discarded items per cell	38	0	0
Maximum number of discarded items per cell	1234	918	1009
Modal frequency of unique raw material types per cell	34	3	5
Minimum unique raw material types per cell	5	0	0
Maximum unique raw material types per cell	64	11	18
Maximum attained raw material transfer distance (grid cells)	69	46	73
Modal transfer distance (grid cells)	3	3	5
Median transfer distance (grid cells)	6	6	8

4.2.2. Transfer distances

Raw material transport ultimately determines most aspects of the lithic discard record but, as noted by Brantingham (2006), archaeological sites only provide a direct record of transfer or displacement distances, on the basis of which transport patterns must be modeled. Such modeling is beyond the scope of this paper, however, and the following analyses focus exclusively on transfer distance patterns evidenced in the outputs of the three simulations. These transfer patterns can be approached from at least two different perspectives, namely: a) the spatial distribution of raw materials procured from a given source, and b) the contribution of different sources to individual assemblages (e.g. Luedtke 1992: 117; Féblot-Augustins 2009). Since each of these perspectives is expected to yield different insights, they are adopted here individually.

4.2.2.1 *The spatial distribution and overall abundance of individual raw material types*

The quantity of any given raw material type is, in all three discard records, highest at, or in cells adjacent to, its source, and rapidly decays with increasing distance from that source (Figure 6a; see also Figure 11 in Brantingham 2003). The actual number of items discarded at their source is highly variable, as is the total contribution of any one source to the overall record: in simulation A, for example, the total quantities of discarded raw materials procured from individual sources ranged between a minimum of 8058 and a maximum of 55997 items (median = 32970; IQR = 11268).

This variability is a consequence of two different factors, namely agent mobility (here a simple random walk) and the local density of sources. Both of these influence the degree to which individual sources are exploited through the same fundamental mechanism: they each partially determine the quantity of materials already present in the toolkits during source encounters, and consequently the amount of fresh materials that agents may collect (see Brantingham 2003). In essence, source densities determine the distances between sources, while agent mobility determines the speed (i.e. number of time-steps) at which those distances can be traversed. These two variables are expected to show opposing trends in terms of their influence on source exploitation since, all else being equal, speed is less important when distances are short.

Simulation data support this observation, as the local density of sources, considered here in terms of the mean distance between a source and its nearest four neighbours[15], was found to be positively correlated with the degree of source exploitation in all three discard records, albeit to different degrees. Not surprisingly, the strongest correlation was observed in simulation A (R_s = .83, n=5000, $p < 2*10^{-16}$; see Figure 6b), where the density of sources was highest, while the weakest correlation was observed in simulation B (R_s = .356, n=500, $p < 2*10^{-16}$), where the density of sources was lowest (for otherwise identical model parameters). The intermediate value of the correlation coefficient obtained for simulation C (R_s = .487, $p < 2*10^{-16}$) can be explained by the higher carrying capacity of the agents (200

[15] The choice of four nearest neighbours is arbitrary, and is used here only to illustrate certain patterns. The optimal number of nearest neighbours to use will depend on the relationship between the global density of sources, the size of the toolkit, and the specified discard rates. Mean distances are used because a fixed number of neighbours is considered, and considering median distances would result in outliers being essentially ignored which, in this particular context, is not desirable.

instead of 100 raw material units), which effectively decreased the distance between sources with respect to the relative depletion rates of toolkits.

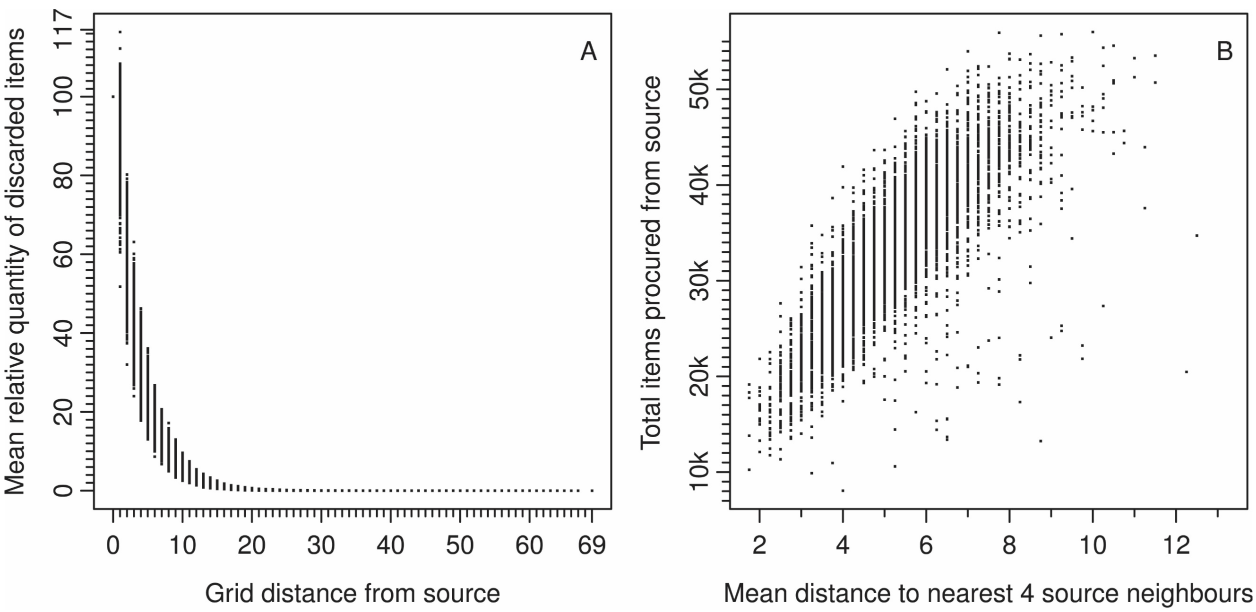

Fig. 6 The relative quantities of materials contributed by individual sources, expressed as a percentage of the number of items discarded at their source, decrease rapidly with increasing distances following a remarkably stable decay curve (*a*). These quantities are expressed as percentages because the absolute number of items contributed by individual sources is correlated with the distances between the sources and their nearest neighbours; subplot *b* illustrates this correlation using mean distances to the nearest four neighbours. All data are from simulation A

Figure 7 shows the distribution of two types of materials (a' and b') from simulation A in the areas surrounding their sources. These sources were selected based on their extreme values for mean distance to the nearest four neighbours, namely 13.25 and 1.75 grid cells respectively. As expected given the preceding discussion, the source of material a' contributed considerably more items to the discard record (54843 items in total, with a maximum of 518 per cell) than did the source of material b' (19150 items in total, with a maximum of 130 items per cell). Interestingly, however, the relative abundance of items at various distances from their sources seems unaffected by the distribution of other sources in the area, being more or less constant in all directions and quantitatively similar at both locations. In both cases, for example, the contour line representing a decay of 50 percent relative to the abundance of materials discarded at their source is more or less circular, and located at about 3 grid cells from the respective sources[16]. The overall effects of this decay pattern (see also Figure 6a) are visible in all three discard records (Figure 5a1, b1, c1), but are particularly evident in those of simulation B and C.

[16] The pattern is more irregular around source b' simply because of the lower sample sizes.

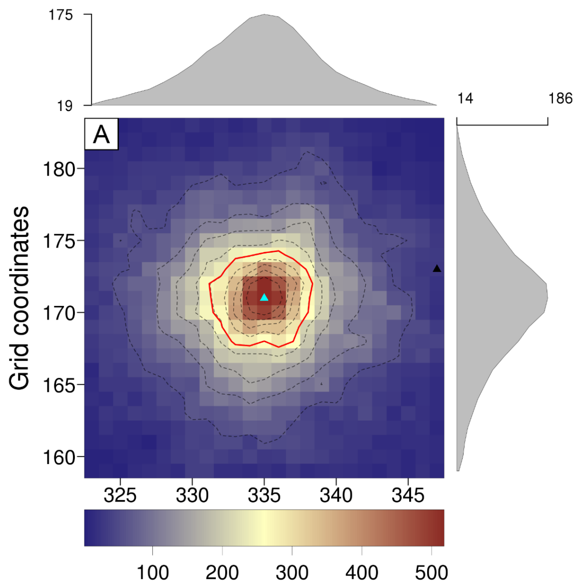

Fig. 7 Spatial distribution of two types of materials from simulation A in the vicinity of their sources (light-blue triangles), which were selected on the basis of their extreme mean nearest (4) neighbour distance values of 13.25 (a), and 1.75 (b) grid cells. Both materials are encountered most frequently at or near their sources, and their frequency decreases at more or less the same rate in all directions regardless of the presence of other sources in the area (black triangles). This is illustrated by the contour lines (dashed gray), which indicate decreases in the quantities of materials in increments of 10 percent relative to the amounts discarded at the sources; in both cases a decrease of 50 percent is observed at a distance of about grid cells (red contour lines). Note, however, that the actual amount of items contributed by the two sources varies dramatically

The relative stability of this decay curve suggests that, unlike the degree to which sources are exploited, the relationship between the relative abundance of a raw material type in a given cell and the distance to its source is virtually unaffected by global source densities or simulation lengths. To test this, the relative abundance of raw materials at a distance of exactly four grid cells from their source was plotted for all three simulations, as well as simulation A truncated to the first 100 million time-steps. The results, shown in Figure 8, confirm this hypothesis: the medians and interquartile ranges are virtually identical in simulation A and B despite the order of magnitude difference in source densities (0.02 versus 0.002) and the considerable (2.5x) difference in simulation lengths between the truncated and full simulation A data set. However, Figure 8 also shows that the parameters of raw material diffusion from sources are strongly affected by the toolkit/discard rate (T/D) ratios, as simulation C, with a T/D ratio of 200, showed a considerably higher median (though similar interquartile ranges) than the other two simulations.

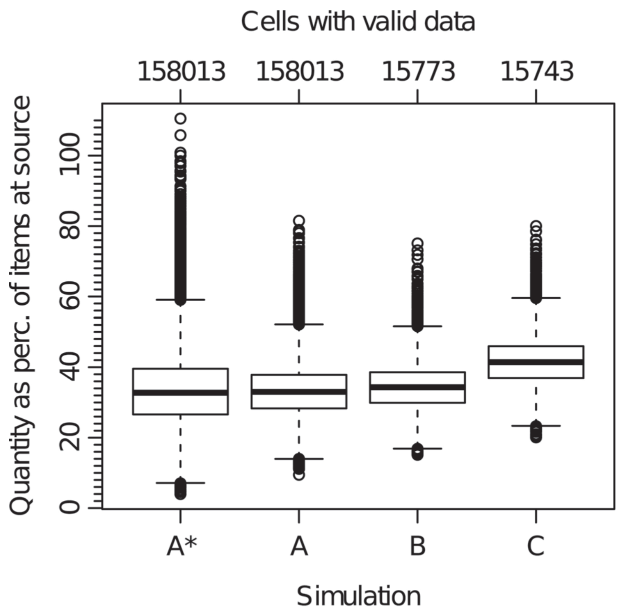

Fig. 8 Relative contribution of individual sources to the discard records at a distance of 4 grid cells. Data from all three simulations, as well as simulation A truncated to a length of 100 million time-steps (denoted here as simulation A*) is included. Note that quantities are expressed here as a percentage of the number of items discarded at their source

4.2.2.2 ***Transfer distances recorded at individual sites (grid locations)***

The frequencies of transfer distances observed in the overall records of the three simulations (Figure 9) are qualitatively similar to those I observed in the toolkits states (Figure S8 in ESM1) and discard records (Figure 3) of the 500 simulations of the replicated model (cf. Figure 9 in Brantingham 2003). The modal and median transport distances evidenced in simulation A and B are identical at 3 and 6 grid cells respectively and, when sampled only at raw material sources, the records show main modes at zero and secondary modes that, with simulations A and B, are identical or nearly identical to the values obtained for all cells. While the T/D ratio does have an effect on the shape of the frequency distributions, the shape seems to be, as suggested by Brantingham (2003: 499) and my own analyses of toolkit patterns (ESM 1), only minimally affected by the environmental density of sources.

These are aggregated statistics, however, and there is reason to suspect that, due to the dependence of cell contents on the spatial configuration of sources in the area, they may not be representative of the patterns evidenced in individual cells. Indeed, if median, modal, and maximal transfer distance values recorded in individual cells are considered, an altogether different picture emerges (Figure 10). Cells from simulations B and C tend to show similar median and modal values, despite differences in T/D ratios, and in both cases these values are considerably higher than in simulation A, in which the density of sources was ten times greater.

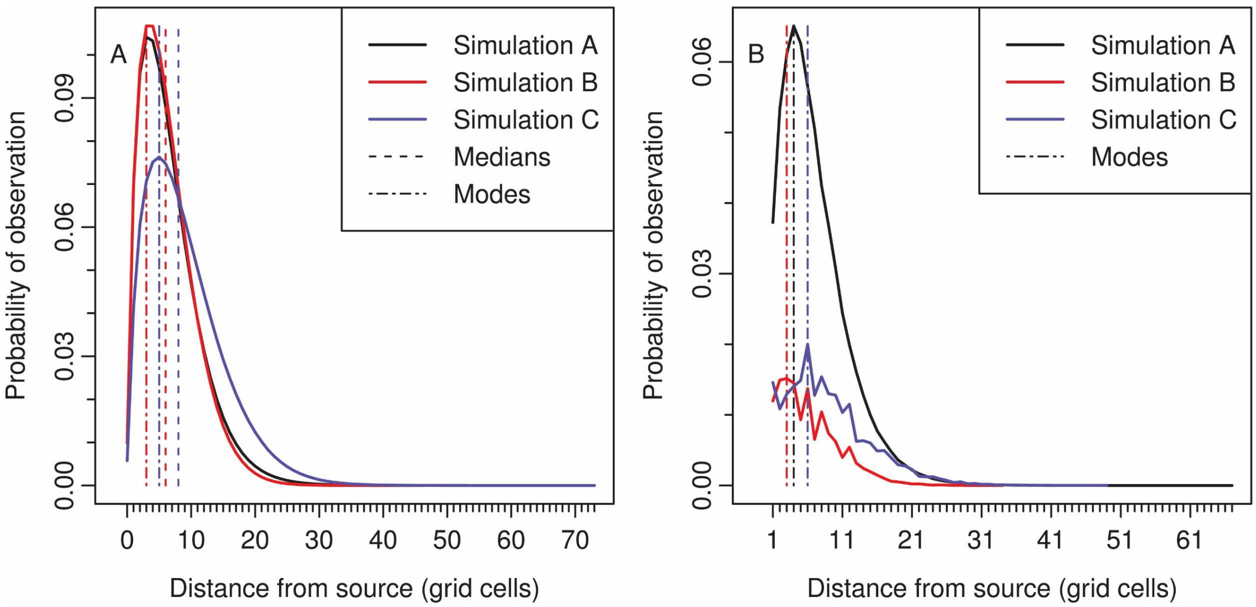

Fig. 9 Probabilities of observing specific transport distances in the aggregated data from all three simulations at: a) all grid locations, and b) at raw material sources

These patterns do not seem to be influenced by simulation length, as the values obtained for the full record of simulation A are virtually indistinguishable from those obtained for the same simulation truncated to a length of 100 million time-steps (Figure 10ab). Maximum transfer distances, on the other hand, do seem to be affected by source densities, T/D ratios, and to a lesser extent simulation lengths (Figure 10c), with a two-fold increase in T/D ratios (simulation C) having approximately the same effect as a ten-fold increase in source densities (simulation A).

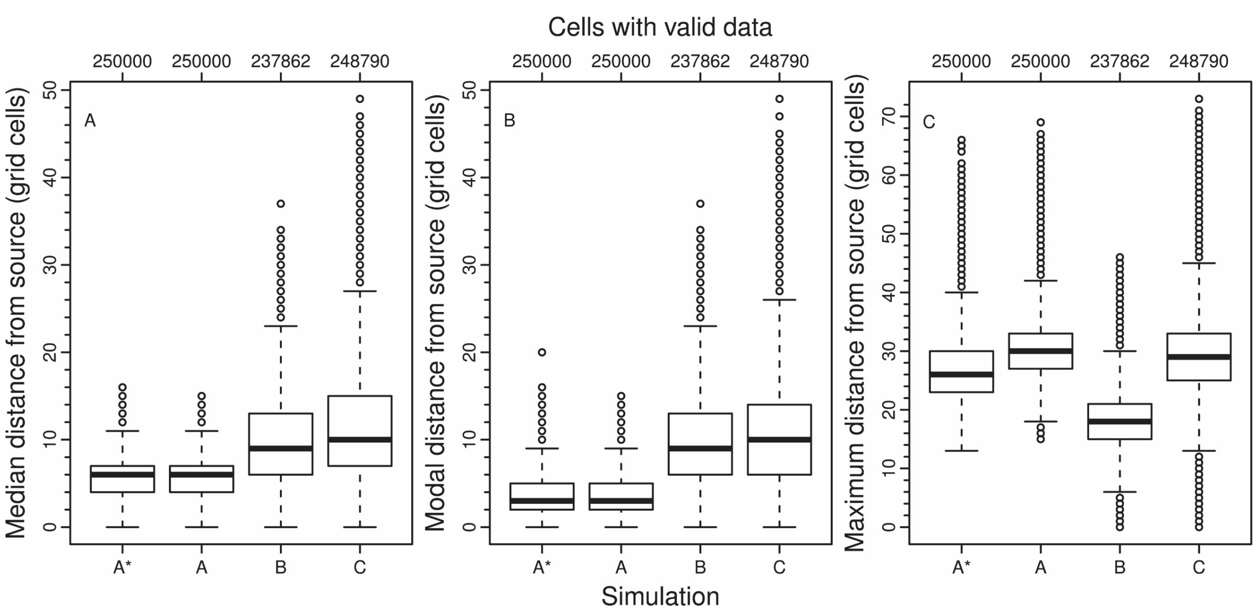

Fig. 10 Median (a), modal (b), and maximum (c) transfer distances observed in the three simulations, as well as simulation A truncated to a length of 100 million time-steps (denoted here as simulation A*)

In short, median and modal transfer distances recorded at individual discard locations do change with different environmental source densities (c.f. Brantingham 2003: 499) and, all else being equal, so do the maximum transfer distances. Consequently, the median ratio of maximal to modal transfer distances is highly variable across the simulations (medians of 9 [IQR=8.76], 1.9 [IQR=1.76], and 3 [IQR=2.75] for simulations A, B, and C respectively), rendering specific yet generally applicable predictions for one distribution parameter based on another virtually impossible to make (cf. Brantingham 2003: 501).

More importantly, perhaps, there is no evidence of a clear and consistent internal mode at individual raw material sources (cf. Brantingham 2003: 499). Indeed, at raw material sources, as well as at other grid locations, the frequency distribution of items procured from various distances is quite variable (Figure 11). This does not mean, however, that such variability is random; in the case of high source density environments at least, most of it is expected to be a function of the spatial configuration of sources in the area (Figure 6b).

To ascertain the degree to which this is so, the variability in transfer distance frequency distributions at one of the ten randomly selected cells shown in Figure 11 (all from simulation A, and all located on raw material sources) was investigated through the discard records produced by an additional nine full simulations (250 million time-steps each). The placement of sources in these simulations was identical to that of simulation A, as were all other general model parameters[17].

As shown in Figure 12, the frequency distribution of materials transferred across various distances at the investigated location is quite stable across multiple simulations, indicating that, as expected, random factors (i.e. the number of times specific paths are traversed) play only a minor role in determining compositional variability, at least with larger assemblages. The relatively low degree of variability observed in the results of the full-length simulations also suggests that it is possible to model the composition of actual archaeological assemblages, and therefore quantify departures from neutral expectations, with a fair degree of precision and known error ranges.

[17] The initial placement of the agent, as well as its mobility and discard behaviours, were fully and independently random in each simulation, as per the model's specifications.

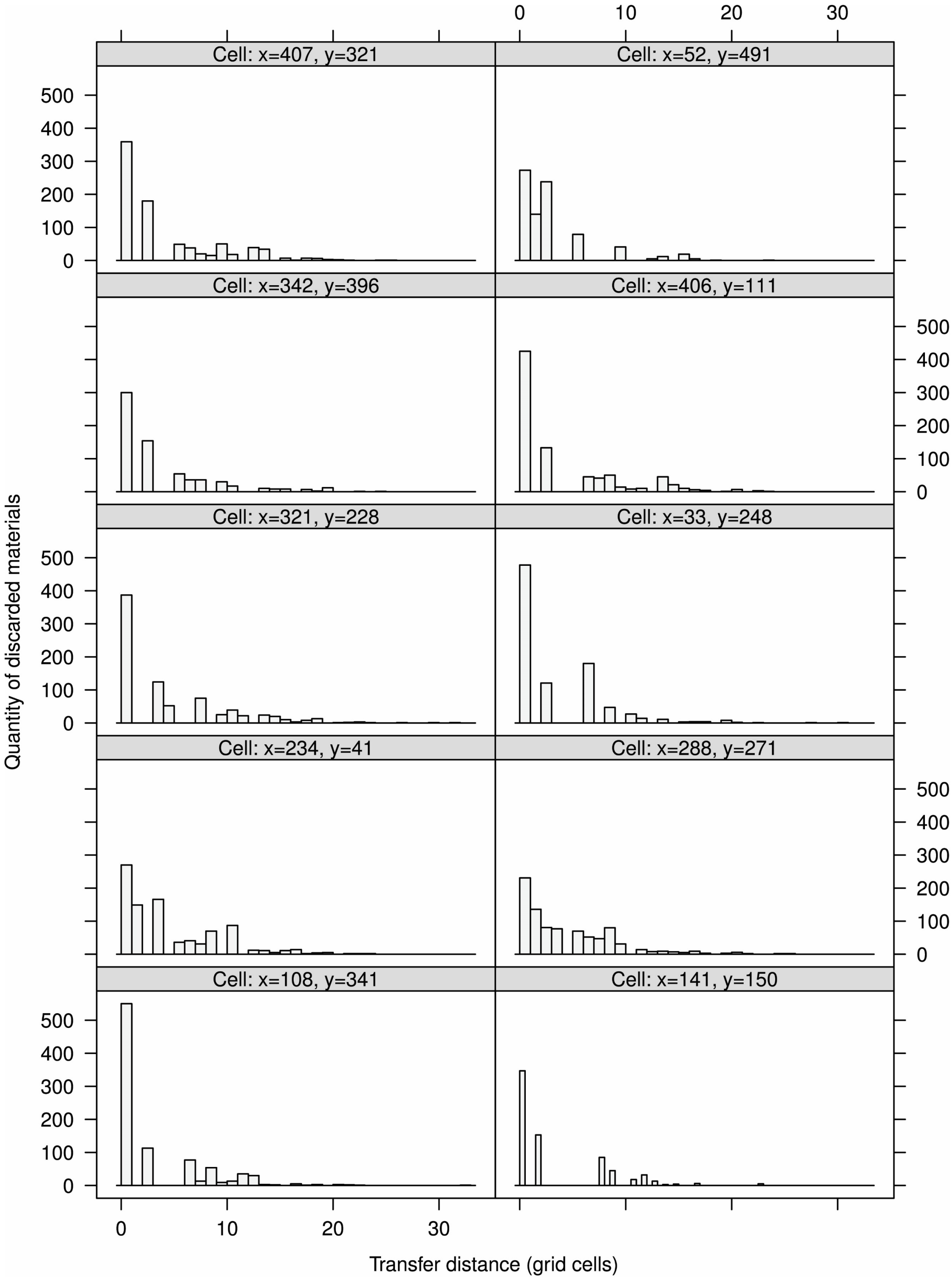

Fig. 11 Frequencies of raw material transfer distances recorded in a random sample of 10 cells containing raw material sources

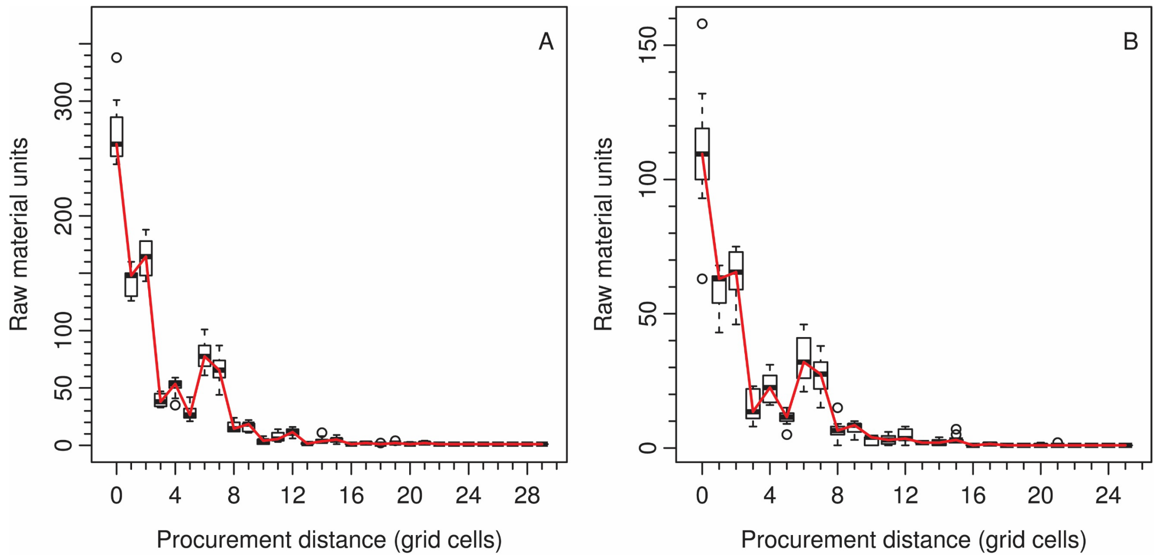

Fig. 12 Variability in the quantity of items procured from various distances at a single grid location, as evidenced in the outputs of 10 simulations run with identical parameters. The location was sampled after 250 million time-steps (a), as well as after only 100 million time-steps (b). Overall trends are illustrated by the line connecting the boxplot medians

4.2.3. Raw material richness

As shown in Figure S6a (ESM 1), the frequency distribution of *toolkit* richness states follows a smooth decline curve from a modal value of zero unique raw material types up to a maximum that depends, partially at least, on the density of sources as well as the toolkit size/discard rate ratios (see also Brantingham 2003: 494). The discard records investigated here demonstrate that all simulation parameters, including simulation lengths, affect not only the maxima and minima of that distribution, but also its overall shape (cf. Brantingham 2003: 494) which, when not truncated by factors such as very low source densities (relative to toolkit size/discard rate ratios) or short simulation lengths, tends to approach normality (Figure 13). They also show, in any case, that the distribution parameters, including minima and maxima, are quite different from those observed in the toolkits (cf. Figure S6 in ESM1).

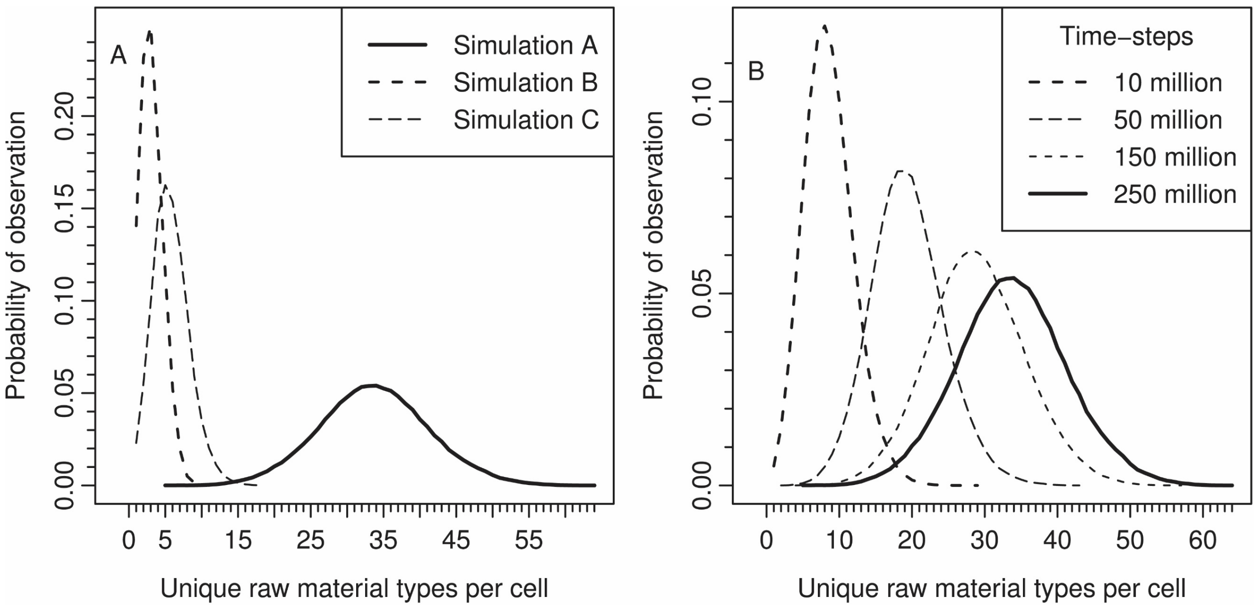

Fig. 13 Probabilities of observing different richness states (i.e. number of unique raw material types) at individual locations across the three simulations (a), as well as simulation A truncated to different lengths (b)

The maximum richness states were observed in simulation A, where the maximum number of unique raw material types represented in a single cell was 64, and the modal value was 34 (see Table 1). These values, which are much higher than those predicted by Brantingham for identical model parameters, do not represent absolute limits: all else being equal, longer simulations will tend to yield even higher median, modal and maximal values (see Figure 13b). Brantingham's observation regarding the maximum attainable toolkit richness states cannot, therefore, be extrapolated to the discard (i.e. archaeological) record. This is simply because, as discussed in section 4.1 above (see, in particular, Figure 2), the processes operating in that record differ from those governing the composition of toolkits.

On the basis of his analyses of toolkit states Brantingham (2003: 501) concluded that "[t]he fact that some raw materials present in the environment are never procured, others are only rarely procured, while a few are commonly procured need not imply raw material selectivity." The discard records analyzed here do not support this conclusion, insofar as "environment" is defined in a reasonable manner. On the contrary, these records show that *all* sources within a certain radius, the size of which depends on simulation length and the toolkit/discard rate (T/D) ratio, are represented at all landscape locations. In simulation A, for example, all sources within a linear radius of 10 grid cells are always represented after an average of 1000 visits per cell (Figure 14), matching, and in most cases exceeding, Brantingham's (2003: 495) predicted foraging radius of 10, itself an overestimation (see ESM1). In fact, only at a radial distance of 17 grid cells does the degree of source representation at any given location drop below an average of 99 percent[18]. It should be noted that with a lower average number of visits per cell (i.e. shorter simulations) this radius of full source representation is smaller, extending to only 8

[18] Consider that, in order to attain a modal representation of 34 unique raw material types per cell, as is the case in the discard record of simulation A, effective exploitation areas of roughly 1700 grid cells (i.e. a linear radius of approximately 23 grid cells) are required at the specified resource density of 0.02 (i.e. 5000 sources on a 500x500 grid).

and 2 grid cells with an average of 600 (150 million time-steps) and 200 (50 million time-steps) cell visits respectively. Still, even in simulations where the maximum attainable assemblage size is approximately 200 items (i.e. 50 million time-steps), over 99 percent of sources are represented, on average, within a linear radius of 9 grid cells.

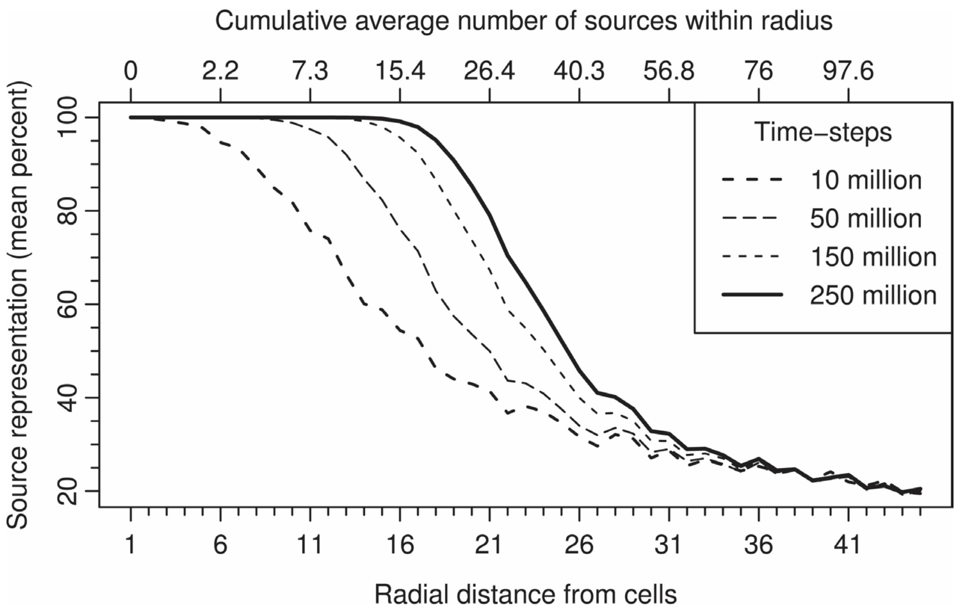

Fig. 14 Mean degree of representation of sources located at different linear distances from cells in simulation A. All 250,000 grid cells were sampled after 10, 50, 150, and 250 million simulation time-steps. Source representation was computed as the number of sources available within a radius of *i-1, i* (where *i* represents the indicated radial distance) of a given cell divided by the number of sources actually represented in that cell and multiplied by 100

Finally, Brantingham (2003: 502) noted that the neutral model predicts a dependence of richness states (i.e. number of unique raw material types represented in an assemblage) on assemblage size. The correlation between the number of unique raw material types and total cell contents is weak in the overall record produced by the simulations (R_s = 0.341, n= 250000, $p < 2*10^{-16}$ for simulation A), but subsamples of individual assemblages do show a moderate correlation between the variables. In the case of the largest cell (i.e. “site”) generated in simulation A, for example, 30 sequential snapshots of approximately 8.3 million time-steps each evidence an apparently linear relationship between sample size and raw material richness (R_s = .67, $p < .001$), and it is likely that this relationship would be stronger still with larger sample sizes; more importantly, perhaps, it is very likely that with increasing sample sizes raw material richness would increase non-linearly, as suggested by the patterns evidenced in Figure 13b and Figure 14. However, any process that would impose fuzzy, steadily decreasing limits on raw material transfer would generate a similar relationship (i.e. dependence of raw material richness on the log of sample size).

5. Discussion

Brantingham (2003) focused on three aspects of raw material variability evidenced in archaeological contexts, namely: a) the number of different sources that are represented in a given assemblage relative to its size, b) the distances over which materials were transported, and c) the frequencies of different material types in relation to procurement distances. On the basis of his analyses he derived a series of predictions about what assemblage variability should look like under neutral behavioural assumptions, and argued, through two case studies, that these predictions seem to hold.

The analyses presented in this paper call into question some of Brantingham's original predictions. They also demonstrate that qualitative similarities in patterns generated by the model and those observed archaeologically cannot be used to prove, or disprove, the adaptive or functional significance of raw material variability, insofar as it can be accepted that both behaviourally neutral factors (e.g. the natural density of sources) *and* selection may have been at play.

The revised model presented here allows for new predictions to be made, based on the influence of relatively isolated variables (e.g. random mobility). Each of these predictions can be evaluated quantitatively, since it has been shown that the composition of individual assemblages can be modelled with a known degree of precision (Figure 12). Interestingly, the rejection of individual predictions may be traceable to behaviourally significant biases in particular domains. Thus, rejecting one prediction may indicate a preference for a specific material, while the rejection of another prediction may indicate biased mobility.

While predictions can be evaluated with a high degree of precision, and it is argued here that, except in extreme cases, such high precision is required in order to detect both the presence and absence of behaviourally meaningful deviations, the model is highly sensitive to any inaccuracies or omissions present in the source distribution and exploitation datasets. In other words, the model may only be useful in cases where highly reliable sourcing studies have been conducted.

Finally, the revised model also provides some insight on topics of wider archaeological relevance. Indeed, the model presented here is also, by necessity, a model for the formation of the archaeological record, although only from the perspective of the lithic components. Consequently, it enables an objective exploration of topics such as the relative importance of stone tools in past adaptations, and of the meaning of archaeological sites. While no conclusions can be reached on the basis of the model, it nevertheless allows for a value-free starting point for discussion of topics of fundamental importance for archaeological research. The foregoing issues are discussed in detail below.

5.1. Brantingham's original observations in the context of the revised model

5.1.1 Assemblage raw material richness

Brantingham (2003: 501) stated that under neutral assumptions raw material richness is always "constrained to be much less than the maximum richness theoretically attainable". If reasonable definitions of theoretical maximum are used[19] (e.g. the number of sources available within the foraging

[19] This theoretical maximum is ill-defined by Brantingham: it cannot be considered in terms of assemblage size, which Brantingham seems to imply, because the number of discarded items found in the assemblage may exceed the number of physically available sources, and it cannot be considered in terms of the number of physically available sources without limiting inquiry to a given area, and it is unclear how large that area should be.

radius), it can be shown that, on the contrary, the modeled behaviours would result in the full representation of *all* available sources in assemblages of sufficient size (Figure 14). Indeed, assemblages generated under the revised model are characterized by substantially higher raw material richness than that observed by Brantingham under identical model parameters (Figure 13). Thus, the analyses presented herein do not support the view that "[t]he neutral model anticipates the low observed richness relative to the environmental density of sources" (Brantingham 2003: 502) at Grotte Vaufrey.

My revised model does support Brantingham's observation that raw material richness co-varies with assemblage size, but does not support the view that biases in procurement probabilities arise from the path dependence of random walks. In fact, the effect of such path dependence can be shown to be minor because, over time, the influence of individual random paths averages out. Rather, procurement probabilities depend on the relationship between toolkit parameters and the specific spatial configuration of the sources in a given area (see Figure 6b).

5.1.2. <u>Transfer distances</u>

Under the original model the "frequency distribution of raw material transport distances [is expected to show] an internal mode" (Brantingham 2003: 501), which is virtually unaffected by the density of sources or toolkit parameters. Brantingham argued that the frequency distribution is also expected to show maximum values that are "three to four times the distance represented by the internal mode" (Brantingham 2003: 501). Analyses of individual assemblages generated by the revised model indicate that resource densities do affect the value of the internal mode (Figure 10), which is neither stable nor easily discernible in many cases (Figure 11). They also show that resource densities and toolkit parameters affect the ratio of modal to maximum transfer distance, which is in general higher than suggested by Brantingham (see Figure 10 and associated discussion). On the other hand the "long right skew" observed by Brantingham (2003: 498) for these distributions was also noted in the output of the revised model (Figure 9), and it is clear that he was right in stating that maximum transport distances do not relate to geographic ranges (Brantingham 2003: 501).

5.1.3. <u>Distance decay patterns</u>

Brantingham noted that the quantity of a given material will "generally [follow] an exponential decline with increasing distance from source" (Brantingham 2003: 501). He argued that materials from distant sources are expected to be found consistently in low quantities, while materials from nearby sources are expected be more common and show high variance. Analyses of the revised model largely support these predictions (Figure 6a).

5.1.4. <u>Summary</u>

The patterns evidenced in the discard records discussed in this paper are inconsistent with some of Brantingham's observations and predictions. Overall, the general similarities in the nature of the patterns generated by the neutral model and those observed in at least some archaeological cases do indicate that inferring specific behaviours on the basis of simple correlations or general shapes of frequency distributions may be unwarranted. However, it seems equally unwarranted to infer, on the same basis, that non-modelled behaviours (i.e. raw material selectivity) played no role in shaping raw material variability. Since we are not faced with an either/or scenario, it is theoretically possible, and even likely, that both modeled (e.g. the natural density of sources) *and* non-modeled (e.g. raw material selection) factors influenced assemblage composition.

As shown in the preceding sections, qualitative similarities in general trends of the kind discussed by Brantingham (2003) are not sufficient to either reject or accept the neutral model, except perhaps in extreme cases. For example, the discard records analyzed here show that patterns of individual cells do not necessarily conform well to aggregated patterns (Figures 9-10). Thus for individual sites (i.e. grid cells) where the predicted frequencies differ significantly from the general patterns discussed by Brantingham, a strong agreement with the latter would in fact indicate that non-modelled behaviours were at play.

Similarly, the general nature of the relationship between assemblage size and raw material richness cannot indicate, by itself, whether the neutral model is to be accepted or rejected - contrary to what is suggested by Brantingham (2003: 501), choice may be assumed to have played a role even if a single source predicted to have been exploited is found to have not been. In short, the neutral model can only be rejected or accepted on the basis of specific, quantitative predictions with known error ranges. Otherwise, the most parsimonious explanation for similarities with the model's output may simply be that the model is not sensitive enough to detect differences.

5.2. Predictions of the revised model

A series of predictions can be made on the basis of the revised model, and deviations from each of these may be traceable to biases in particular behavioural domains (e.g. mobility, raw material preferences, and carrying capacities or discard behaviours). These predictions are outlined below in a very general manner so as to convey the overall expectations. However, they can, and indeed *must* be, evaluated quantitatively through the variability evidenced in the output of multiple simulations, each run with an identical placement of sources and an identical set of toolkit/mobility parameters.

5.2.1. <u>Distribution of raw materials around a source</u>

Raw materials from any given source are expected to *always* be found in similar quantities at locations with similar access costs, regardless of access direction. For example, with the model parameters employed in simulation A, it would be expected that, if 100 units of raw material a' were discarded at their source, two assemblages located at 4 grid cells from that source would, in 50% of cases, contain between 28 and 38 units of this material (see Figure 8).

Under conditions of equal toolkit size/discard rate (T/D) ratios, major deviations from these expected abundances (i.e. an assemblage with 200 units of raw material a' at a distance of 4 grid cells from the source a') can only be explained by behaviours that affect mobility patterns. This is because the relative abundance of a certain material at a given distance from its source is minimally affected by simulation lengths and unaffected by source densities (Figures 7,8). Moreover, a preference or avoidance of the source would be manifested through its under- or over-exploitation.

5.2.2. Degree of source exploitation

Under the modeled behaviours it is expected that the most heavily utilized raw materials will be those from relatively isolated sources (Figures 6b, 7). In cases where the quantity of raw materials from this source follows a similar decline curve in all directions, which would be consistent with random mobility and similar accessibility from all sides, deviations from this expectation likely reflect factors that resulted in an avoidance of that source or raw material. Preferential exploitation of such a source would be expected to manifest in terms of a higher than predicted, but never lower, abundance of raw materials.

5.2.3. Assemblage size as a function of distance to nearby sources

Given the sharp decline in raw material abundances with increasing distance from their source (Figure 6a, Figure 7), it is expected that large sites will only form at or very near to raw material sources. The effects of this are most visible in the discard records of simulations B and C (Figure 5b1,c1). Deviations from this expectation can only be explained by behaviours that result in biased, non-random movement.

5.2.4. Raw material frequencies in an assemblage

All else being equal, the probability of observing a given number of raw material units of a specific type in an assemblage will depend not only on the distance between the assemblage and the raw material source, but also on the distance between the source and its neighbours (Figure 6b). Thus, the relative contributions of two sources located at an equal distance from a given assemblage are only expected to be equal if the distribution of sources in the landscape is perfectly uniform. In cases where the evidenced variability is consistent with random mobility, significant deviations from modelled frequencies are likely indicative of factors that resulted in preferential procurement or avoidance of particular raw materials.

5.2.5. Maximum transfer distances as a function of the density of sources

All else being equal, it is also expected that maximum transfer distances, as well as the ratio of maximum to median or modal transport distances observed in individual assemblages will be smaller under conditions of low source densities than under conditions of high source densities (Figures 9, 10). Deviations from these expectations may reflect biased mobility patterns, although they may also reflect other factors such as lithic recycling.

5.2.6. Raw material richness as a function of the density of sources

Finally, the number of unique raw material types represented in individual assemblages is expected to be low under conditions of low source densities (Figure 13). However, regardless of the local density of sources, all raw material sources are expected to be represented within a given radius, the size of which co-varies with assemblage size (Figure 14). Deviations from this expectation likely reflect a preference or avoidance of specific raw materials.

5.2.7 Summary

As shown in the preceding sections (see Figure 12 and associated discussion), assemblage composition can be modeled with a high and known degree of precision. The various expectations discussed above show that the revised model is, theoretically at least, capable of detecting the influence of both modeled and non-modeled behaviours quantitatively and within well-defined error ranges. The attainable precision does not guarantee accuracy, however, and overall the model is expected to be very sensitive to incorrect parameters. The availability of high-quality sourcing data is therefore essential for accurate interpretations of the model's output. Indeed, depending on its proximity to an assemblage, omitting a single raw material source may substantially alter the outcome of a simulation, and so would an incorrect assessment of the frequency with which individual sources are represented.

A further limitation of the model pertains to the fact that we have no direct way of estimating toolkit size/discard rate (T/D) ratios and these have a major impact on the model's output (Figures 5, 8, 10). While such estimation is beyond the immediate scope of this paper, at least two tentative solutions can be proposed: one approach would be to use maximum transport distances as proxies (Brantingham 2006), while another approach would be to estimate T/D ratios on the basis of the degree to which a subsample of assemblages (e.g. different levels at one site) fit the output of the model when different ratios are employed. Both of these approaches are problematic, and this subject will be treated in detail in a future paper. However, the advantage of the second approach is that, if different assemblages are found to be best explained by different T/D ratios, we can at least conclude that the assemblages evidence different behaviours (e.g. either a different capacity to carry stone or different rates of discard).

5.3. The wider implications of the revised neutral model for archaeological research

Regardless of whether practical or theoretical problems (e.g. estimation of T/D ratios) prevent the composition of individual archaeological assemblages from being accurately modeled, the revised model presented here provides some critical insights of wider relevance to the discipline. Amongst these, perhaps the most important is the fact that, as hinted by Brantingham (2003) and demonstrated here, the common practice of computing expected frequencies from distances to source is flawed, regardless of how such distances are measured (e.g. 'as the crow flies' or taking into account terrain difficulty). This is because the degree to which sources are exploited depends, to a rather large extent (a variation of over half an order of magnitude was observed in simulation A), on their proximity to other sources (Figure 6b). Interestingly, Browne and Wilson (2011) report a significant correlation between raw material utilization and inter-source distances at the Bau de l'Aubesier. The reported correlation was weak, and the authors eventually dropped inter-source distances from their model, but it is likely that a stronger correlation would have been observed if distances to multiple nearest neighbours had been used (see discussion associated with Figure 6b).

In any case, the analyses presented here also show that assemblages do not reflect, as it is all too often assumed (e.g. Brantingham 2003) the average composition of toolkits, even if all tools are assumed to have had equal use-lives. Moreover, they show that lithic densities, which in the context of the model are equivalent to “site” sizes (all grid cells are of the same dimension), are, by themselves, poor predictors of occupational histories. Indeed, even in cases where the influence of technologies, curation strategies, and sedimentation/erosion rates can be accounted for, major differences in assemblage growth rates can be observed (see section 4.2.1).

Ultimately, the model presented here is, as already pointed out, a model for the formation of the archaeological record; in this regard it is an imperfect model, of course, since only lithic components are considered, and a large number of relevant factors (e.g. topography) are not included. Nevertheless, it offers a glimpse of what an archaeological record *undisturbed* by factors such as erosion and differential sedimentation rates should look like under the simplest behavioural scenario. As such, it offers a value-free starting point for dialogue on the meaning of concepts such as “archaeological site” which are of fundamental importance to archaeological research. In the context of the model such sites are nothing more than random aggregates of items discarded over time by highly mobile agents which do not, in any meaningful sense of the word, dwell at any one location. While real sites may well represent episodes of continued habitation, the implications of the simple alternative suggested here are far too important to ignore, particularly in cases where the overall patterns generated by the model resemble archaeologically observed ones well.

As pointed out by Brantingham (2003: 503-504), it is essential to refine the model by including new variables and studying their effects on the overall patterns. Aside from Brantingham’s (2003) own suggestions, it would be useful to study the effects of lithic recycling behaviours, since these may result in sites themselves being treated as raw material sources. It would also be useful to model semi-random mobility patterns, whereby agents gravitate around either raw material sources or sites, moving randomly until a certain condition is reached (e.g. toolkits become dangerously depleted, or too much simulation time has been spent away from a site), and then directly towards a target (i.e. site or source). Such mobility behaviours could be used to investigate whether archaeological variability reflects a need for constant access to lithic resources or landscape locations (i.e. sites), since preliminary analyses suggest that patterns generated under such conditions can be differentiated from those discussed in this paper with relative ease. What the foregoing analyses show, in any case, is that the model presented here can be a useful exploratory tool. As such, it allows not only for the objective assessment of the effects of different parameters and variables, but also for the detection of new variables that may be of interest (e.g. spatial distribution of sources at a local scale).

6. Conclusions

The analyses presented in this paper have shown that while Brantingham’s (2003) neutral model correctly simulates raw material procurement and transport behaviours (i.e. the overall toolkit patterns were successfully replicated), it stops short of modeling how such behaviours translate into archaeologically visible patterns. Insofar as it can be accepted that discard records are a more direct proxy for the archaeological record than toolkits, my analyses failed to support some of Brantingham’s original predictions. More importantly, however, I have shown that qualitative similarities between predicted and observed patterns are not sufficient to reject, or accept, the neutral model.

I introduced a revised model aimed at overcoming some of the original limitations. Detailed analyses of the discard records produced under this model have shown that, if parameters such as the toolkit size/discard rate ratio can be estimated, the raw material composition of archaeological assemblages can be predicted with a known and relatively high degree of precision. It is therefore possible, in principle at least, to detect deviations from model predictions even in those cases where factors such as the natural density of sources played a dominant role. More importantly, however, the ability to make quantitative predictions with known error ranges implies that the model can be rejected, or accepted, in a mathematically sound way.

I have argued that the attainable precision may be misleading if the underlying provenance dataset is inaccurate or incomplete. Indeed, the model is expected to be highly sensitive to: a) the omission of sources actually exploited in the past, and b) the erroneous identification of such sources in the record. In this regard the model may prove quite useful in determining the exact impact of such errors on the interpretation of raw material variability - to the best of my knowledge, the effects of different errors in sourcing datasets has yet to be evaluated quantitatively.

Overall, this paper has shown that models of the kind discussed here have much to contribute not only to raw material sourcing studies, but also to the wider discipline of archaeology, since they are, by definition, also models for the formation of the archaeological record. Indeed, they allow for value-free, objective discussions on topics of fundamental importance to archaeological research, including the meaning of archaeological sites, the nature of the archaeological record, and the relative importance of lithics in the past.

Acknowledgements

I would like to express my gratitude to Will Archer, Sam Lin, Marcel Weiß, Dawit Desta, Harold Dibble, and Shannon McPherron for their insightful comments on earlier drafts of this manuscript, and to Mike Richards for the encouragement and for making this research possible.

References

Adams, B., & Blades, B. S. (Eds.). (2009). *Lithic Materials and Paleolithic Societies*. Oxford, UK: Wiley-Blackwell.

Amick, D. S. (2007). Investigating the behavioral causes and archaeological effects of lithic recycling. In S. P. McPherron (Ed.), *Tools versus Cores: Alternative Approaches to Stone Tool Analysis* (pp. 223-252). Newcastle, UK: Cambridge Scholars Publishing.

Andrefsky, W. J. (1994). Raw-Material Availability and the Organization of Technology. *American Antiquity, 59*(1), 21-34.

Andrefsky, W. J. (2009). The Analysis of Stone Tool Procurement, Production, and Maintenance. *Journal of Archaeological Research, 17*, 65-103.

Arakawa, F. (2006). *Lithic raw material procurement and the social landscape in the Central Mesa Verde region, A.D. 600-1300*. PhD thesis, Washington State University,

Baddeley, A., & Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software, 12*(6), 1-42.

Binford, L. R. (1979). Organization and Formation Processes: Looking at Curated Technologies. *Journal of Anthropological Research, 35*(3), 255-273.

Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd edition ed.). New York: Springer.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *PNAS, 99*(suppl. 3), 7280-7287.

Brantingham, P. J. (2003). A Neutral Model of Stone Raw Material Procurement. *American Antiquity, 68*(3), 487-509.

Brantingham, P. J. (2006). Measuring Forager Mobility. *Current Anthropology, 47*(3), 435-459.

Braun, D. R., Plummer, T., Ditchfield, P., Ferraro, J. V., Bishop, D. M. L. C., & Potts, R. (2008). Oldowan behavior and raw material transport: perspectives from the Kanjera Formation. *Journal of Archaeological Science, 35*, 2329-2345.

Braun, D. R., Plummer, T., Ferraro, J. V., Ditchfield, P., & Bishop, L. C. (2009). Raw material quality and Oldowan hominin toolstone preferences: evidence from Kanjera South, Kenya. *Journal of Archaeological Science, 36*, 1605-1614.

Braun, D. R., Rogers, M. J., Harris, J. W., & Walker, S. J. (2010). Quantifying Variation in Landscape-Scale Behaviors: The Oldowan from Koobi Fora. In S. J. Lycett, & P. R. Chauhan (Eds.), *New Perspectives on Old Stones: Analytical Approaches to Paleolithic Technologies* (pp. 167-182). New York: Springer.

Browne, C. L., & Wilson, L. (2011). Resource selection of lithic raw materials in the Middle Palaeolithic in southern France. *Journal of Human Evolution, 61*, 597-608.

Clarkson, C. (2007). *Lithics in the Land of the Lightning Brothers: The Archaeology of Wardaman Country, Northern Territory* (Terra Australis, Vol. 25). Canberra: ANU E Press.

Clarkson, C., & Bellas, A. (2014). Mapping stone: using GIS spatial modelling to predict lithic source zones. *Journal of Archaeological Science, 46*, 324-333.

Cole, S. C. (2002). *Lithic Raw Material Exploitation Between 30,000 BP and 40,000 BP in the Perigord, France*. PhD thesis, University of Washington.

Dibble, H. L. (1991). Local Raw Material Exploitation and its Effects on Lower and Middle Paleolithic Assemblage Variability. In A. Montet-White, & S. Holen (Eds.), *Raw Material Economies Among Prehistoric Hunter-Gatherers* (University of Kansas Publications in Anthropology, Vol. 19). Lawrence: University of Kansas.

Dowle, M., Short, T., Lianoglou, S., Srinivasan, A., & Antonyan, w. c. f. R. S. a. E. (2014). data.table: Extension of data.frame. R package version 1.9.4. http://CRAN.R-project.org/package=data.table.

Driscoll, K. (2010). *Understanding quartz technology in early prehistoric Ireland*. PhD thesis, University College Dublin.

Duke, C., & Steele, J. (2010). Geology and lithic procurement in Upper Palaeolithic Europe: a weights-of-evidence based GIS model of lithic resource potential. *Journal of Archaeological Science*, 37(4), 813-824).

Elseberg, J., Magnenat, S., Siegwart, R., & Nüchter, A. (2012). Comparison of nearest-neighbor-search strategies and implementations for efficient shape registration. *Journal of Software Engineering for Robotics, 3*(1), 2-12.

Enquist, B. J., Sanderson, J., & Weiser, M. D. (2002). Modeling Macroscopic Patterns in Ecology. *Science, 295*(5561), 1835-1837.

Féblot-Augustins, J. (1993). Mobility Strategies in the Late Middle Palaeolithic of Central Europe and Western Europe: Elements of Stability and Variability. *Journal of Anthropological Archaeology, 12*, 211-265.

Féblot-Augustins, J. (1999). La mobilité des groupes paléolithiques. *Bulletins et Mémoires de la Société d'anthropologie de Paris, NS, 11*(3-4), 219-260.

Féblot-Augustins, J. (2009). Revisiting European Upper Paleolithic Raw Material Transfers: The Demise of the Cultural Ecological Paradigm? In B. Adams, & B. S. Blades (Eds.), *Lithic Materials and Paleolithic Societies* (pp. 25-46). Oxford: Wiley-Blackwell.

Fernandes, P., Raynal, J.-P., & Moncel, M.-H. (2008). Middle Palaeolithic raw material gathering territories and human mobility in the southern Massif Central, France: first results from a petro-archaeological study on flint. *Journal of Archaeological Science, 35*, 2357-2370.

Gotelli, N. J., & McGill, B. J. (2006). Null versus neutral models: what's the difference? *Ecography, 29*(5), 793-800.

Gould, R. A. (1978). The Anthropology of Human Residue. *American Anthropologist, 80*, 815-835.

Gould, R. A., & Saggers, S. (1985). Lithic Procurement in Central Australia: A Closer Look at Binford's Idea of Embeddedness in Archaeology. *American Antiquity, 50*(1), 117-136.

Hijmans, R. J., & Jacob van Etten (2012). raster: Geographic analysis and modeling with raster data. R package version 2.3-33. http://CRAN.R-project.org/package=raster.

Janssen, M. A. (2012). Introduction to Agent-Based Modeling. Retrieved from https://www.openabm.org/book/introduction-agent-based-modeling

Janssen, M. A., & Oestmo, S. (2013). A Neutral Model of Stone Raw Material Procurement. (version 1): CoMSES Computational Model Library. https://www.openabm.org/model/3957/

Kuhn, S. L. (1991). "Unpacking" Reduction: Lithic Raw Material Economy in the Mousterian of West-Central Italy. *Journal of Anthropological Archaeology, 10*, 76-106.

Lamigueiro, O. P., & Hijmans, R. (2014). rasterVis. R package version 0.35. https://cran.r-project.org/package=rasterVis

Lefebvre, H. (2007). *The Production of Space* (D. Nicholson-Smith, Trans.). Malden, MA: Blackwell Publishing.

Lengyel, G. (2015). Lithic raw material procurement at Bodrogkeresztúr – Henye Gravettian site, northeast Hungary. *Quaternary International, 359-360*, 292-303.

Luedtke, B. E. (1992). *An Archaeologist's Guide to Chert and Flint* (Archaeological Research Tools, Vol. 7). Los Angeles: UCLA Institute of Archaeology.

Matsumoto, M., & Nishimura, T. (1998). Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. *ACM Transactions on Modeling and Computer Simulation, 8*(1), 3-30.

McCall, G. S. (2012). Ethnoarchaeology and the Organization of Lithic Technology. *Journal of Archaeological Research, 20*, 157-203.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed., Monographs on Statistics and Applied Probability, Vol. 37). London: Chapman and Hall.

McPherron, S., & Dibble, H. L. (1999). The lithic assemblages from Pech de l'Azé. *Préhistoire Européenne, 15*, 9-43.

Moutsiou, T. (2011). *The Obsidian Evidence for the Scale of Social Life during the Palaeolithic*. PhD thesis, Royal Holloway, University of London.

Myers, A. M. (1986). *The organizational and structural dimensions of hunter-gatherer lithic technology: theoretical perspectives from ethnography and ethnoarchaeology applied to the Mesolithic of mainland Britain with a case study from Northern England*. PhD, University of Sheffield.

Oestmo, S., Janssen, M. A., & Marean, C. (2014). *The Effect of Spatial Clustering on Stone Raw Material Procurement.* Paper presented at the Advances in Computational Social Science and Social Simulation, Barcelona: Spain.

Pearce, E., & Moutsiou, T. (2014). Using obsidian transfer distances to explore social network maintenance in late Pleistocene hunter-gatherers. *Journal of Anthropological Archaeology, 36*, 12-20.

Pearson, S. M., & Gardner, R. H. (1997). Neutral Models: Useful Tools for Understanding Landscape Patterns. In J. A. Bissonette (Ed.), *Wildlife and Landscape Ecology: Effects of Pattern and Scale* (pp. 215-230). New York, NY: Springer-Verlag.

Perreault, C., & Brantingham, P. J. (2011). Mobility-driven cultural transmission along the forager-collector continuum. *Journal of Archaeological Science, 30*, 62-68.

Pop, C. M. (2013). *Raw material properties and edge attrition rates: A high-resolution study of unretouched experimental flakes.* Poster presented at the 5th International Symposium on Chert and Other Knappable Materials, Iaşi, Romania, 20-24 August.

R Core Team (2014). R: A Language and environment for statistical computing. (3.1.2 ed.): R Foundation for Statistical Computing.

RStudio (2014). RStudio. (0.98.1102 ed.). Boston, MA: RStudio Inc.

Sillitoe, P., & Hardy, K. (2003). Living Lithics: ethnoarchaeology in Highland Papua New Guinea. *Antiquity, 77*(297), 555-566.

Spinapolice, E. E. (2012). Raw material economy in Salento (Apulia, Italy): new perspectives on Neanderthal mobility patterns. *Journal of Archaeological Science, 39*, 680-689.

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering, 13*, 22-30.

Woods, A. D. (2011). *The effects of lithic raw material quality on Aurignacian blade production at Abri Cellier*. PhD thesis, University of Iowa.

Xie, Y. (2013). testit: A simple package for testing R packages. R package version 0.3. https://cran.r-project.org/package=testit

Yamada, M., & Ono, A. (Eds.). (2014). *Lithic Raw Material Exploitation and Circulation in Prehistory: A Comparative Perspective in Diverse Palaeoenvironments* (ERAUL, Vol. 138). Liège, Belgium: University of Liège.

Electronic Supplementary Material 1:

An independent replication of Brantingham's neutral model - summary of results

1. Simulation specifications:

Five hundred simulations of the replicated model were run with the baseline parameters specified by Brantingham (2003) and summarized in Table S1; all raw material source and initial agent locations were generated randomly for each of simulation. Unless otherwise noted, all results described below are based on data generated by these. These data consist of the following information, recorded at each time-step:

a. The position of the agent
b. The contents of its toolkit
c. The distance between the agent and the closest raw material source
d. The number of raw material units that were procured, if a source was encountered during the time-step.
e. Information on the type and toolkit position of the raw material item removed from the toolkit, if an item was discarded during the time-step.

Table S1 List of core model variables and baseline simulation settings.

Variable	**Variable description**	**Units**	**Baseline setting**
GridSize	Side length of the simulated square grid	Grid cells ($\mathbb{N}$)	500
ForagerLocation	Forager location as a grid coordinate	Coordinate 2-tuple ($\mathbb{N}_0$, $\mathbb{N}_0$)	Random = (0-499,0-499)
NrRawMatSources	Number of unique raw material sources	$\mathbb{N}$	5,000
RMSourceLocations	Location of raw material sources	Coordinate 2-tuple ($\mathbb{N}_0$, $\mathbb{N}_0$)	Random = [(0-499,0-499), 0-499,0-499), …]
ToolKitSize	Maximum <ToolKit> size	$\mathbb{N}$	100
ToolKitDepletionRate	Raw material consumption rate	$\mathbb{N}$	1
ToolKit	List of coordinate tuples for each raw material source in the toolkit	Coordinate 2-tuple ($\mathbb{N}_0$, $\mathbb{N}_0$)	Empty list
NrSteps	Simulation time step number (tick)	$\mathbb{N}_0$	0
MoveLength	Length of move	Grid cells ($\mathbb{N}$)	1
MoveProbability	Probability of moving to any of the cells in the Moore neighbourhood or staying in place	$\mathbb{R}$	1/9

All random variables drawn from uniform probability distributions

2. Data validation:

The output of all simulations was validated after these had finished running but prior to analysis, in order to ensure data integrity and full compliance with the definitions of Brantingham's (2003) neutral model. The validation tests (ca. 36 individual tests in total, depending on the characteristics of the output) are summarized in Table S2; all tests were passed by the simulations discussed here.

Table 2 Summary of validation tests conducted on simulation outputs.

Output component	Test description
Integrity of utilized raw material sources data	The source file utilized by the simulation contains the requested number of sources.
	All entries in the utilized source file are located within the bounds of the simulated environment, and are all unique.
Integrity of toolkit state data	The data file contains the correct number (and types) of variables. Time-step data are sequential (incremented by one).
Integrity of discard record data	The number of input events in the discard record matches the number of discard events recorded in the toolkit state data. The timing of the input events matches the timing of discard events in the toolkit states data, and the items entering the discard record are identical to those discarded from the toolkit
	The coordinates in the discard record data match the location of the agent as recorded in the toolkit state data for all discarded items.
Agent positions	Changes in the position of the agent along incremental time-steps are consistent with movements in Moore neighbourhoods (i.e. the agent only moves to adjacent cells).
	There is no significant difference between agent position changes and movement probabilities derived from a uniform random distribution with 9 possible outcomes (chi-square test). The uniformity of the distribution of p values across all simulations was checked.
Distances between the agent and the nearest sources	At least one source listed in the source output file is located at the recorded distance from the agent at each time-step, and no source is found in closer proximity to the agent.
Raw material procurement	Procurement events are recorded only at time-steps where the agent's position matches that of one of the raw material sources listed in the source output file.
	The number of procured raw material units (the toolkit size at the current time-step minus the toolkit size at the previous time-step) is equal to the specified maximum toolkit size minus the number of items currently in the toolkit
	The toolkit is always full at time-steps when the agent's position matches the location of a raw material source listed in the source output file.
	Only one source, which corresponds to the source indicated by the agent's current position, is represented in the procured materials.
Raw material discard	No discard event recorded in time-steps where toolkit is empty.
	Item(s) are always discarded at time-steps when the toolkit is not empty.
	The number of discarded items matches specified discard rates.
	The toolkit size recorded in the discard information variables matches the size of the toolkit in the previous time-step.
	The type of the discarded raw material unit matches the type of the unit at the index indicated by the discard information variables in the previous time-step.
	The number of units of the discarded type in the previous time-step is the same as the specified discard rate (one), accounting for source encounters.
Discard and Procurement	Sequential differences in toolkit sizes are explainable by procurement and discard behaviours.
Simulation end conditions	The last move attempt would have placed the agent beyond the boundaries of the simulated environment (edge encounter), or the agent moved to the location of an unvisited source after having visited 199 unique sources (source encounter).

Note that the randomness of resource placement, though not tested, conforms to theoretical expectations (Figure S2). All tests were conducted in R

3. **Preliminary observations**:

A number of problems were identified during the re-implementation of Brantingham's model. The first of these pertains to Brantingham's discussion of "foraging areas," which he defines as the "area[s] around a raw material source that can be effectively exploited with material from that source [and are] thus defined by radius *d*" (Brantingham 2003: 495; italics in original). This concept of foraging area is problematic in the context of a model which uses square tiling and allows for diagonal movements without penalty because the area defined by the radius is square, not circular (i.e. a forager can move diagonally over 10 grid cells in 10 time steps). Consequently, in the context of his model an area of radius *d*=10 (20x20 cells) will measure 400 grid cells and will contain an average of 8 raw material sources at a resource density of 0.02 sources per grid cell (i.e. 5000 sources on a 500x500 grid); conversely, if linear distances are considered, the area defined by that radius would be 314.16 grid cells, containing on average only 6.2 raw material sources. In short, using linear distances to derive foraging radii and average source densities (e.g. Brantingham 2003: 495-496) is not consistent with the definitions of the model.

To ascertain the actual mean of the distribution of agent displacement distances after 100 moves (i.e. the time it would take for a full toolkit to be cleared if no sources are encountered), which represents, according to Brantingham (2003: 495), the foraging radius *d*, a simple model was run a million times. In the simulations, which occurred on a 500x500 grid, the agent was placed at the center of the grid (x=250, y=250) and allowed to move randomly (1/9 probability of moving to a neighbouring cell or staying put) for 100 time steps. At the end of each simulation the minimum number of time-steps required to return to the point of origin was recorded. The resulting distribution (Figure S1) has a mode of 7, a median of 9, and a mean of 9.213, which is, as expected, somewhat lower than the number (10) used by Brantingham. More importantly, Figure S1 shows not only that the range of possible displacement values is high, but also that the distribution is skewed and, as indicated by the central tendency values listed above, most observations fall below the mean; thus, it could be argued that the modal or most frequently observed value of 7, rather than the mean, should be used as a proxy for the size of the foraging radii. In any case, it is clear that only probabilistic estimates of, rather than absolute values for, foraging radii can be given.

A second problem pertains to Brantingham's calculation of the maximum number of unique raw material types that can be expected in a toolkit. He suggests that this number "can be estimated empirically from the maximum number of unique sources that might be found within a foraging area of radius *d*" (2003: 495). On the one hand his calculations are incorrect because they are based on linear distances which, as discussed above, are not applicable in the context of his model; on the other hand, it must be noted that the maximum resource density in an area of a given size is not fixed - it will vary between simulated environments, and therefore it cannot be predicted in absolute terms *a priori*.

Indeed, there is a non-zero (albeit infinitesimally small) probability that an arbitrarily large area will be fully covered with sources provided that the number of sources is larger than the number of cells in that area, and that the distribution is indeed random (i.e. not uniform).

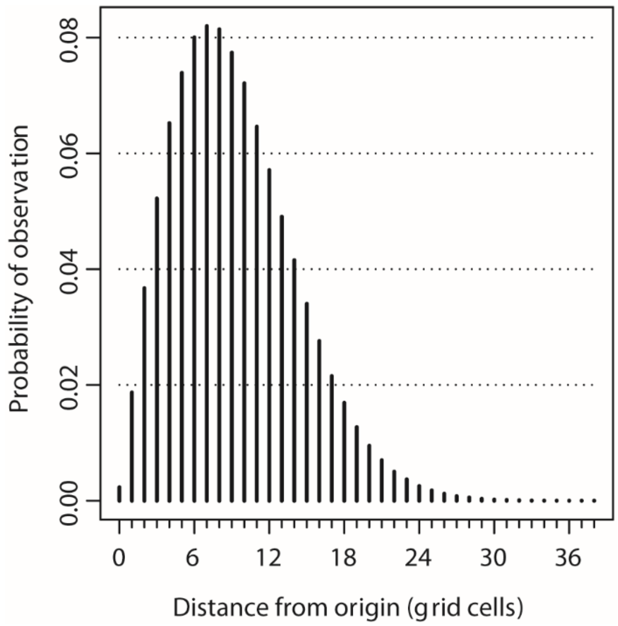

Fig. S1 Probability of traversing a given distance, where distance is defined as the minimum number of steps required to return to the starting location, using a random walk with 1/9 probabilities of moving to cells in the Moore neighborhood or staying in place. The figure summarizes 1 million runs (100 million time-steps)

For any given area the density of randomly distributed sources follows a Poisson distribution. Using Brantingham's displacement figure (*d*=10) and default resource density parameter (0.02 sources per cell), the foraging area in a square grid world would be 400 grid cells, and the λ parameter of the probability distribution would be 8. As shown in Figure S2, the probability of observing resource densities higher than the maximum suggested by Brantingham, namely 11, is relatively high; moreover, as also shown in the figure, the observed distribution of sources across the 500 main simulations run with the same density parameter closely matches this theoretical distribution. Since the density of local clusters cannot be predicted *a priori* except in probabilistic terms, the logic behind Brantingham's (2003: 496) calculations is difficult to follow, given that in the context of a single simulation the actual maximum density of sources does not have to be predicted – it can be computed with perfect accuracy.

A third problem arises from the fact that, if only probabilistic estimates of the foraging areas and maximum source densities can be given, it stands to reason that the maximum richness of toolkits will depend, to a certain degree at least, on the length of the simulations, as longer simulations have a higher chance of producing low likelihood events. As expected, data from the 500 main simulations show a strongly positive and highly significant correlation between these variables (Kendall's $\tau = 0.548$, $p < 2.2e^{-16}$), which suggests that arbitrarily ending the simulations upon encountering 200 unique sources, or upon encounters with the edges of the grid, may have biased Brantingham's analyses. In any case, as calculated by Brantingham, the maximum attainable toolkit richness state is a more or less arbitrary number from the upper tail of the probability distribution, one which represents neither the theoretical maximum nor the expected average maximum that would be observed over a large number of, or longer, simulations. The strong correspondence between the values observed and predicted by Brantingham is therefore due to chance (cf. Table 2 and related discussion in Brantingham 2003: 496).

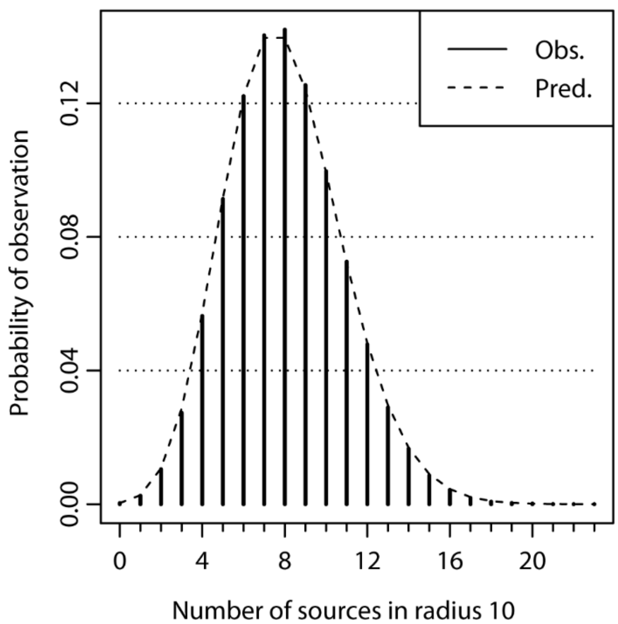

Fig. S2 Predicted and observed probability of a given number of raw material sources being allocated in an area of 400 grid cells, with source densities set to 0.02 (5000 sources randomly placed on a 500 × 500 grid). The observed probabilities were computed from 500 simulated environments, each split into 625 squares of 400 grid cells (*n* = 312,500). Note that the actual probability of observing the average value (8) is only about 14 %

4. Replication results:

4.1. General characteristics of the simulations:

The duration of individual simulations was highly variable, ranging from a minimum of one time-step to a maximum of 50866 (median = 13234, IQR = 27323), for a cumulative total of 8.11 million time-steps. The majority were terminated upon encounters with the edges of the simulated world (n=362, or 72.4%), although longer simulations were predictably terminated more frequently due to encounters with 200 unique raw material sources (Figure S3).

These relatively short simulations allowed the agents to cover only small areas of the grid, the maximum number of unique cells visited by an agent in a single simulation being only 12094 (median=4429.5, IQR=8387), or ca. 4.84% of the total space available to them. One consequence of such low coverage is that, in the context of a random (as opposed to uniform) distribution of resources, which guarantees variability in the clustering of sources at a local scale (Figure S2), conclusions drawn from but a handful of simulations may be unreliable.

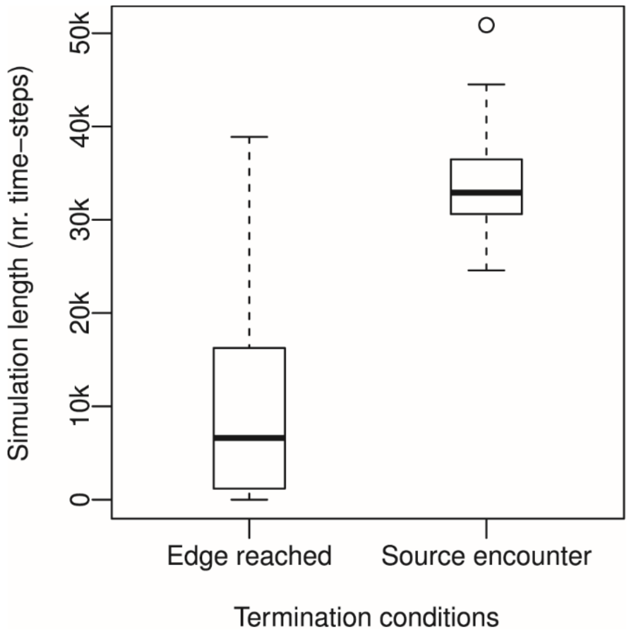

Fig. S3 Distributions of simulation lengths by termination conditions. Data from all 500 simulations is included

In terms of the number of sources visited by the simulated agents, the values varied between a minimum of 0 and a maximum of 969, with a median of 273.5 (IQR=550.25). This is consistent with the number of source encounters reported by Brantingham (2003: 499) for one simulation, namely 513. The agents seldom wandered far from raw material sources, being found most often at a distance of but two to three cells from the one nearest to them, and only very rarely at distances of 10 grid cells or more (Figure S4). While not surprising given the high density of sources - the linear average nearest neighbour distance between these was only 3.61 (SD=0.025)[1] – the ratio between the median number of simulation steps and the median number of visited sources is very high (ca. 48), indicating that source encounters are relatively rare.

[1] Note that this number is higher than would be expected from nearest neighbour analysis (Clark and Evans 1954, 1979), regardless of whether correction factors (e.g. Donnelly 1978) are used or not. This is because such analysis is based on linear distances between neighbours, and cannot be used in the context of Brantingham's model, as sources are allocated on discrete rather than continuous space. In such discrete space the minimum possible distance between two sources is not fixed (i.e. the distance to an adjacent source can be one if the source is at the left, for example, but it will be 1.41 if located at the top-left). Thus, the discrepancy between observed and expected nearest neighbour values is not indicative of a non-random distribution of sources (see Figure S2).

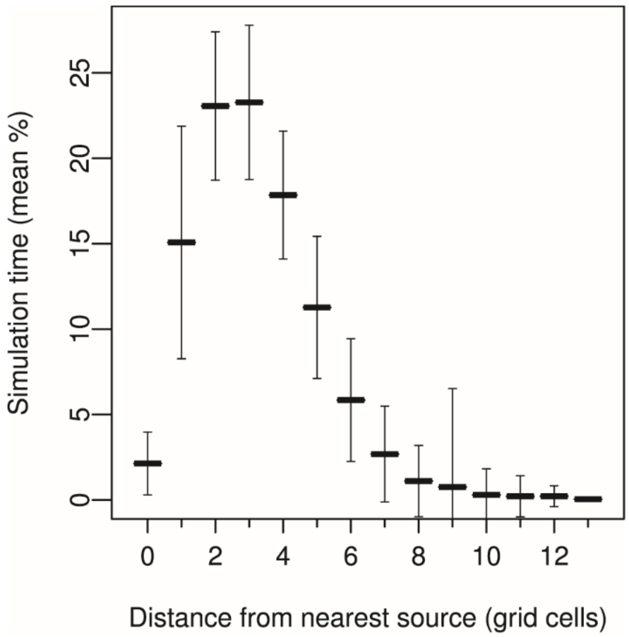

Fig. S4 Frequency of observed distances between the simulated agent and the nearest raw material sources, expressed as the average percentage (±1 SD) of simulation time over all 500 simulations

4.2. Mobile toolkit states:

With respect to the mobile toolkits, these were in an empty state between 0 and 100% of the simulation times, with a median of 35.13 (IQR=8.62) and a mean of 38.32 percent (SD=17.03); if only those simulations lasting 10000 time-steps or more (n=281) are considered, the percentage of simulation time spent in an empty toolkit state ranges from a minimum of 22.73 to a maximum of 44.67 percent, with a median of 34.58 (IQR=5.7) and a mean of 34.65 (SD=4.11) percent. Brantingham (2003: 494) reported that in a representative simulation the toolkit was empty approximately 25 percent of simulation time; while that value falls within the range of variability observed here, it is not representative of the overall results (Figure S5a).

The large number of time-steps spent without access to lithic raw materials could be problematic if one were to assume that such access was critical to forager survival. More important than empty toolkit occurrences, however, are the lengths of periods spent in such a state. The aggregated data from all 500 simulations (Figure S5b) indicates that the lengths of consecutive empty toolkit states were most frequently quite short (<20 time steps), but also that the probability of observing longer sequences is very high (81.2 percent for periods of 20 time-steps or longer, 15.6 percent for periods of 200 time-steps or more); indeed, the average length of empty toolkit periods was 108.75 (median=70), and the maximum recorded length was 1869. These numbers are surprisingly high given the density of sources (see above), and bring into question the survivability of foragers under Brantingham's neutral model[2].

[2] This was noted by Oestmo et al. (2014) also, but they investigated simulations run with non-random, clustered sources. In fact, as shown here, the presence of such clustered sources is not necessary to bring into question the survivability of Brantingham's foragers.

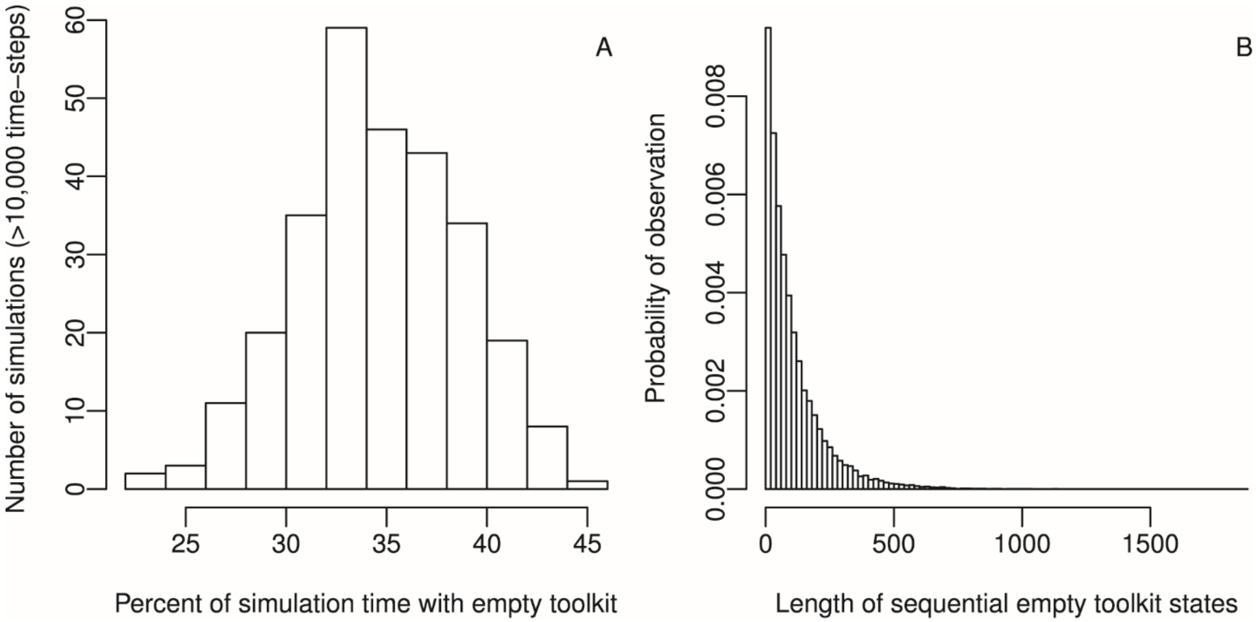

Fig. S5 **a** Frequency of empty toolkit occurrences across 281 simulations lasting 10,000 time-steps or more, expressed as a percentage of simulation time, and **b** the probability of observing sequential empty toolkit states of a given length. Subplot *b* includes all sequential empty toolkit state observations (*n* = 25,923) from all 500 simulations, binned in intervals of 20 time-steps

As expected given these observations, the modal average toolkit richness state observed across all simulations is zero (Figure S6a). From this modal value the distribution of mean richness state frequencies follows a smooth decline curve up to a maximum of 13 unique raw material types, and has a median value of 1 (IQR=2). This distribution is qualitatively similar to that reported by Brantingham for a single simulation (Figure 6 in Brantingham 2003: 494), although as expected given his flawed calculations of maximum attainable richness states (see section S1.3 above) and the uncommonly low frequencies of empty toolkit states reported for the simulations, there are slight differences in the median and maximum values. Overall, most simulations only attained a maximum richness state of 8 unique raw material types (Figure S6b), although the median maximum observed richness was somewhat lower at 7 (IQR = 3).

As noted by Brantingham (2003: 498), the maximum number of unique raw material types increases with the log of the amount of material in the toolkit. As shown in Figure S7, a virtually identical relationship can be observed in the output of the simulations discussed here (compare with Figure 8 in Brantingham 2003: 498); Figure S7 also shows, however, that all combinations of toolkit size and raw material richness can be observed, and that the overall relationship between the two variables is rather different.

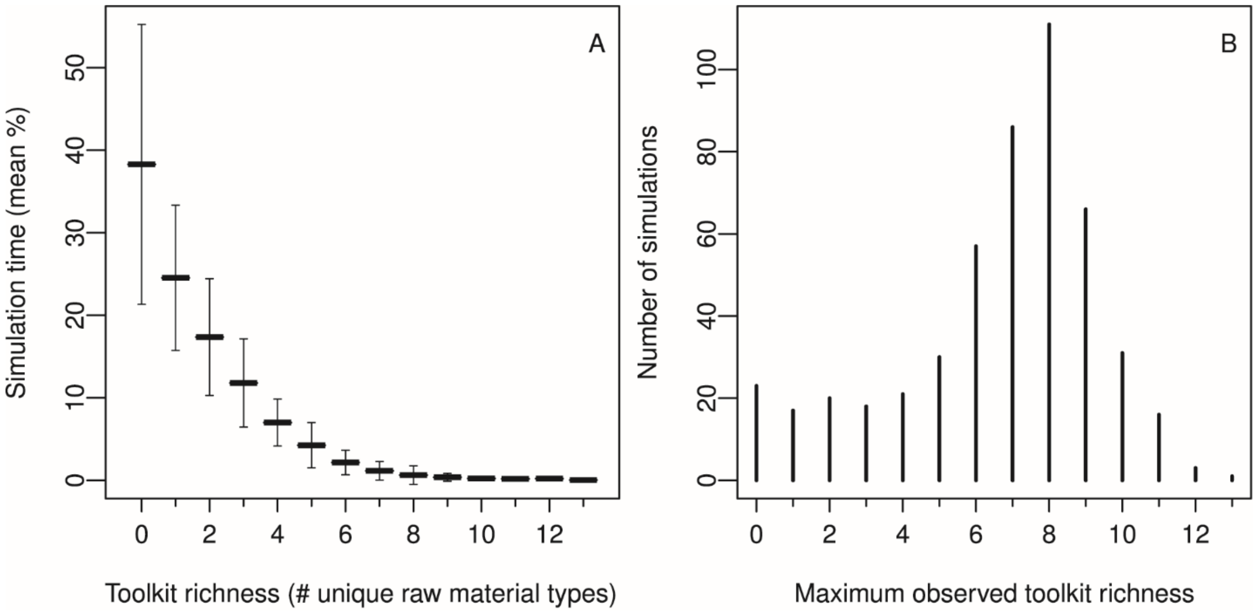

Fig. S6 **a** Average (±1 SD) simulation time spent in different toolkit richness states, and **b** the frequency of maximum attained toolkit richness states across all 500 simulations

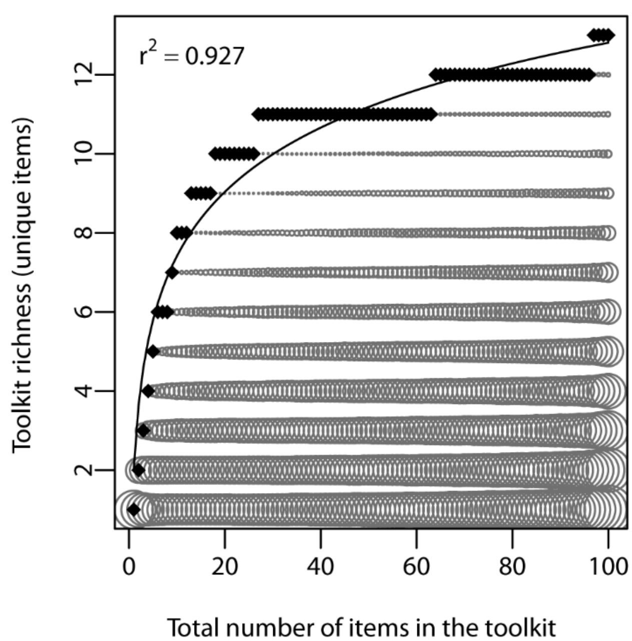

Fig. S7 Relationship between toolkit raw material richness and toolkit size in the aggregated data from all simulations. *Black diamonds* indicate maximum toolkit richness values attained at each toolkit size; as observed by Brantingham (2003: 498), a strong positive correlation ($r^2 = 0.927$, $p < 2.2e^{-16}$) between these maximum richness values and the log of the quantity of materials in the toolkits can be observed (*black regression line*). However, the overall relationship between the two variables is much less clear when all data points are considered ($n = 5.29$ million). These data points are represented here as *circles* whose size increases with the cube root of the number of observations for a given combination, so as to account for the extreme differences (the number of observations per combination ranges from 1 to 50719)

4.3 Transport distances:

The average transport distance patterns that result from combining the outputs of all 500 simulations (Figure S8) are qualitatively very similar to those observed by Brantingham (2003: 499); although the frequency of items found at their source is considerably higher when toolkits are sampled only at encounters with sources (Figure S8b), sample size effects and Brantingham's choice of histogram bin widths likely account for much of the difference. The maximum transport distance of 61 grid cells is also higher than observed by Brantingham (2003: 499), but this is expected given the probabilistic nature of the phenomena under consideration and the much larger number of observations (n=333.7 million versus less than 20000). Importantly, and as predicted by Brantingham (2003: 499), the modal transport distance across all observations (3 grid cells) is identical to the "internal mode" evidenced when only observations at raw material sources are considered.

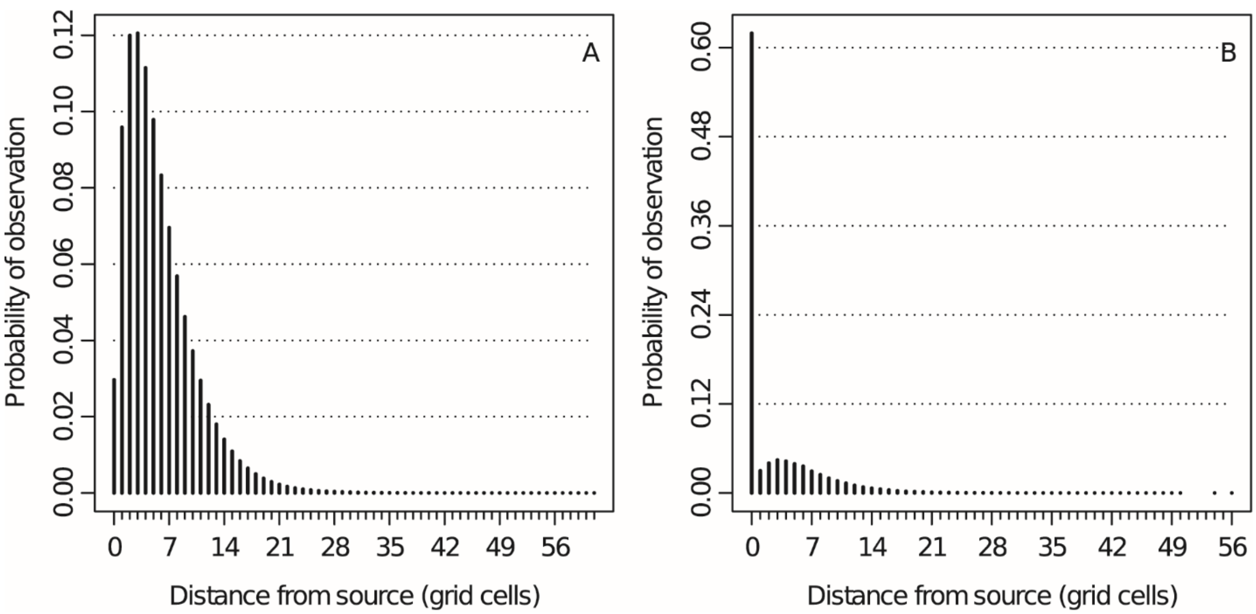

Fig. S8 Probability frequency distribution of raw material transport distances recorded in the mobile toolkits throughout the length of the 500 simulations, when **a** all data points are considered (n = 333.73 million) and **b** when toolkits are sampled at raw material sources (n = 159.83 million)

The pattern of decay in the quantity of a given raw material type with increasing distance from its source (Figure S9) is also qualitatively similar to that reported by Brantingham (cf. Figure 11 in Brantingham 2003); indeed, although the mean values vary somewhat from those reported by him, they fall well within one standard deviation from these.

Some differences were also noted, however, and these bring into question two fundamental relationships predicted by Brantingham. First, Brantingham (2003: 499) predicted that the modal/median transport distances will fall in the interval [d/2,d], where d is the foraging radius (10 for the default simulation parameters); the modal value obtained here falls outside of this range, while the

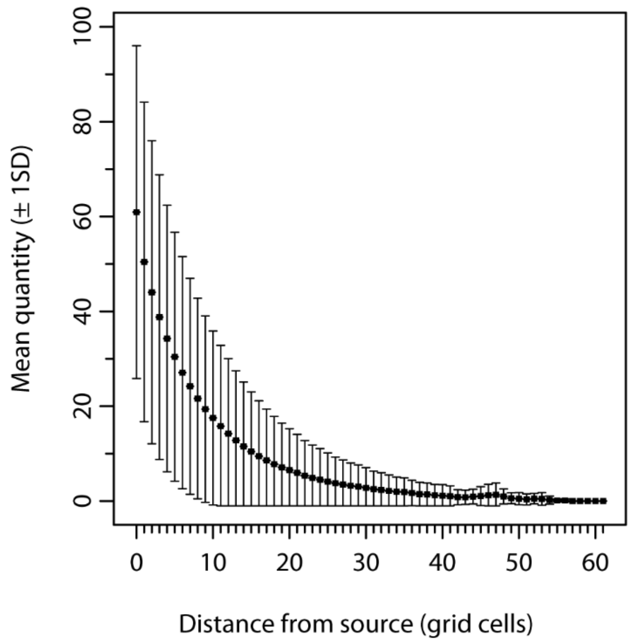

Fig. S9 Quantity of items in the mobile toolkit as a function of distances from their source. Data represent the averaged (±1 SD) frequencies recorded in the 500 simulations

median (5 grid cells) corresponds to its lower limit[3]. Secondly, Brantingham (2003: 500) predicted that the maximum transport distance should be approximately three to four times the radius *d,* although subsequently he states that the maximum distance is expected to be three to four times the mode (Brantingham 2003: 501-506). Notwithstanding this confusion, which has implications for Brantingham's (2003: 501-504) analyses of archaeological parallels[4], it can be noted that the maximum distance observed in the output of the 500 simulations discussed here is over 20 times larger than the mode[5],

[3] Note that the modal and median values reported by Brantingham do fall within the range of variability observed in the outputs of individual simulations; thus two simulations (0.4%) attained a median value of 9 (the maximum observed), and 22 simulations (4.4%) attained modal values equal to, or greater than, 5 (observed maximum = 11). Note also that the overall median value reported here was computed from a random sample (with replacement) of one million data points.

[4] Brantingham (2003: 498-499) reported a modal transport distance of 5 and a maximum of 43 grid cells; using these values, the maximum transport distance is 8.6 times larger than the mode, not 3-4 times as stated by the author. In the archaeological parallels discussed by Brantingham (Grotte Vaufrey and the Middle Palaeolithic record of Central Europe) the reported maximum transport distances are 2 and 4.6-6 times larger than the modes, falling well below the expected values. More importantly, Brantingham reported that the probability of observing a transport distance of 10 grid cells or more, which would be equivalent to twice the mode, is approximately 33 percent (see caption to Figure 10 in Brantingham 2003: 500). Thus, according to Brantingham's predictions, 33 percent of the artifacts recovered from Grotte Vaufrey would be expected to have been transported from a distance that is equivalent to, or greater than, the actual maximum observed. It is also telling, of course, that the maximum transport distances reported for the entire Middle Palaeolithic of Central Europe fall below the predictions of short (i.e. low sample size) simulations.

[5] Note that the probability of observing transport distances greater than 9 and 12 grid cells (i.e. three and four times the modal value observed in the aggregated data from the 500 simulations) are 16.9 and 7.9 percent respectively. In effect, this means that, using the figures obtained here, approximately 8 out of every 100 raw

and approximately 6 times higher than Brantingham's analytically derived foraging radius *d*. In fact, 126 simulations, or 25.2% of the total, yielded maximum transport distances greater than Brantingham's maximum predicted distance of 40 (four times radius *d*).

Be that as it may, Brantingham's observation that "median and modal transport distances do not change appreciably with different environmental raw material densities" (2003: 499) is accurate. An additional 200 simulations, run with source density parameters of 0.002 (500 sources) and 0.06 (15000 sources) respectively confirmed Brantingham's assertion that in cases of low raw material densities the "distribution loses its right skew and clusters more tightly around the [center]" (Brantingham 2003: 499). Across these simulations I observed a shift in modal and median values from 2 to 3, and 4 to 5 respectively, which for all practical purposes can be considered as minor. It should be noted, however, that the increased clustering around the modal and median values does imply that, under conditions of lower source densities, the probability of observing lower maximum transport distances is higher. In other words, the neutral model predicts that, all else being equal, maximum transport distances under conditions of lower raw material source densities should be lower than under conditions of higher source densities.

5. Summary of findings:

The patterns generated by the replicated model discussed here are qualitatively similar to those reported by Brantingham, and all values reported by that author fall within the range of variability evidenced in my simulations. However, his values are seldom representative (e.g. Brantingham's reported median transport distance is higher than the median observed in 99.6% of the 500 simulations), rendering several of Brantingham's conclusions problematic. In particular, I could not replicate the postulated relationship between modal and maximum transport distances, nor that between median/modal transport distances and the size of the foraging radius. The use of methods incompatible with the definitions of the model (e.g. those relying on linear distances) also renders some of Brantingham's analyses problematic; such is the case, for example, with his calculations of maximum toolkit richness states and of the size of foraging areas.

In general the simulation lengths and outputs were highly variable, and the probability of observing representative patterns across all investigated relationships in the context of a single or small number of simulations is low. This is mainly due to the arbitrary termination conditions stipulated by Brantingham, which result in a very low degree of coverage of the simulated environment and therefore in a high susceptibility to local differences in the placement of raw material sources. In this regard it should be pointed out that, although the average number of sources found in an area of radius 10 is 8, the actual probability of observing such an average value is only about 14% (see Figure S2).

The arbitrary end conditions also make it difficult to establish the actual parameters of the various distributions, because very short simulations, virtually always terminated by encounters with the edges of the simulated world (Figure S3), have undue weight. For example, the means and standard deviations

material units (e.g. artifacts) are expected to have been transported over a greater distance than the maximum predicted.

of the percentages of simulation time spent in an empty toolkit state change considerably depending on whether all simulations or only those lasting longer than 10000 time-steps are considered (see section 4.2 above). While it could be argued that only simulations above a certain length should be included in analyses, it is unclear what that threshold should be; at the very least simulations should be allowed to continue until the agent achieves complete or near complete coverage of the grid.

Aside from these issues, and as noted by Oestmo et al. (2014) also, the survivability of foragers under Brantingham's neutral model is questionable, since the simulated agents do not have any access to raw materials for relatively long periods. This is a natural outcome of employing random walk strategies, which cannot guarantee constant access to resources. Consequently, if it can be assumed that people required constant access to lithic raw material sources in the past, the mobility strategy employed by Brantingham's model cannot be used, and either a fully non-random or semi-random (i.e. random movement until certain condition is met) mobility strategy should be modeled instead.

References

Brantingham, P. J. (2003). A Neutral Model of Stone Raw Material Procurement. *American Antiquity, 68*(3), 487-509.

Clark, P. J., and Francis C. Evans (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology, 35*(4), 445-453.

Clark, P. J., and Francis C. Evans (1979). Generalization of a Nearest Neighbor Measure of Dispersion for Use in K Dimensions. *Ecology, 60*(2), 316-317.

Donnelly, K. P. (1978). Simulations to determine the variance and edge effect of total nearest-neighbour distance. In I. Hodder (Ed.), *Simulation Studies in Archaeology* (pp. 91-95). Cambridge: Cambridge University Press.

Oestmo, S., Marco A. Janssen, and Curtis Marean (2014). *The Effect of Spatial Clustering on Stone Raw Material Procurement.* Paper presented at the Advances in Computational Social Science and Social Simulation, Barcelona: Spain.