sociological science

Supplement to:

Greenland, Fiona, and Michelle D. Fabiani. 2021. "Collaborative Practices in Crisis Science: Interdisciplinary Research Challenges and the Syrian War." Sociological Science 8: 455-479.

Table S1. Descriptive statistics (N = 35)

	N	%	Mean	Range
Analysts—All	21	60.00		
Analyst only	13	37.14		
Analyst-Detector	6	17.14		
Analyst-Decision Maker	2	5.71		
Detectors—All	9	25.71		
Detector only	2	5.71		
Detector-Analyst	7	20.00		
Detector-Decision Maker	0	0		
Decision Makers—All	5	14.28		
Decision Maker only	3	8.57		
Decision Maker-Analyst	2	5.71		
Decision Maker-Detector	0	0		
Education	35		2.771	1-3
1. Bachelor's Degree	2	5.71		
2. Master's Degree	4	11.43		
3. Doctorate (incl. JD)	29	82.86		
Distance from Translational Work[1]	35		1.571	1-2
1. Near (2 or fewer steps)	15	42.86		
2. Far (3 or more steps)	20	57.14		
Position in the Conflict Archaeology field[2]	35		1.457	1-2
1. Adjacent	19	54.29		
2. Embedded	16	45.71		
Time commitment[3]	35		2.314	1-3
1. Less than 1 hr/week	3	8.57		
2. Part Time (1-33 hrs/week)	18	51.43		
3. Full Time (<34 hrs per week)	14	40.00		

[1] **Near/Far** is spatial. We use it to refer to a person's distance from the translational research process. Our coding decision *Near* or *Far* is scaled on process steps removed from crisis policy-making. We coded people as Near if they were 2 or fewer steps away, and Far if they were 3 or more steps away. In this context, the mean shows a relatively even split in the sample, with a slight emphasis toward Far.
[2] **Embedded/Adjacent** is temporal. We use it to refer to the amount of focus devoted to conflict archaeology. This variable factors in time commitment (full- or part-time) and the direct relevance of an individual's work to conflict archaeology. For example, individuals could be described as Adjacent if they worked full-time on the satellite remote sensing instruments that collected data for conflict archaeologists, but did not analyze the data themselves. As indicated by the mean (1.457), our sample is relatively evenly split between Embedded and Adjacent, with a slight emphasis towards Adjacent commitment.
[3] On average, over the course of the crisis science response from early 2015 through mid-2017.

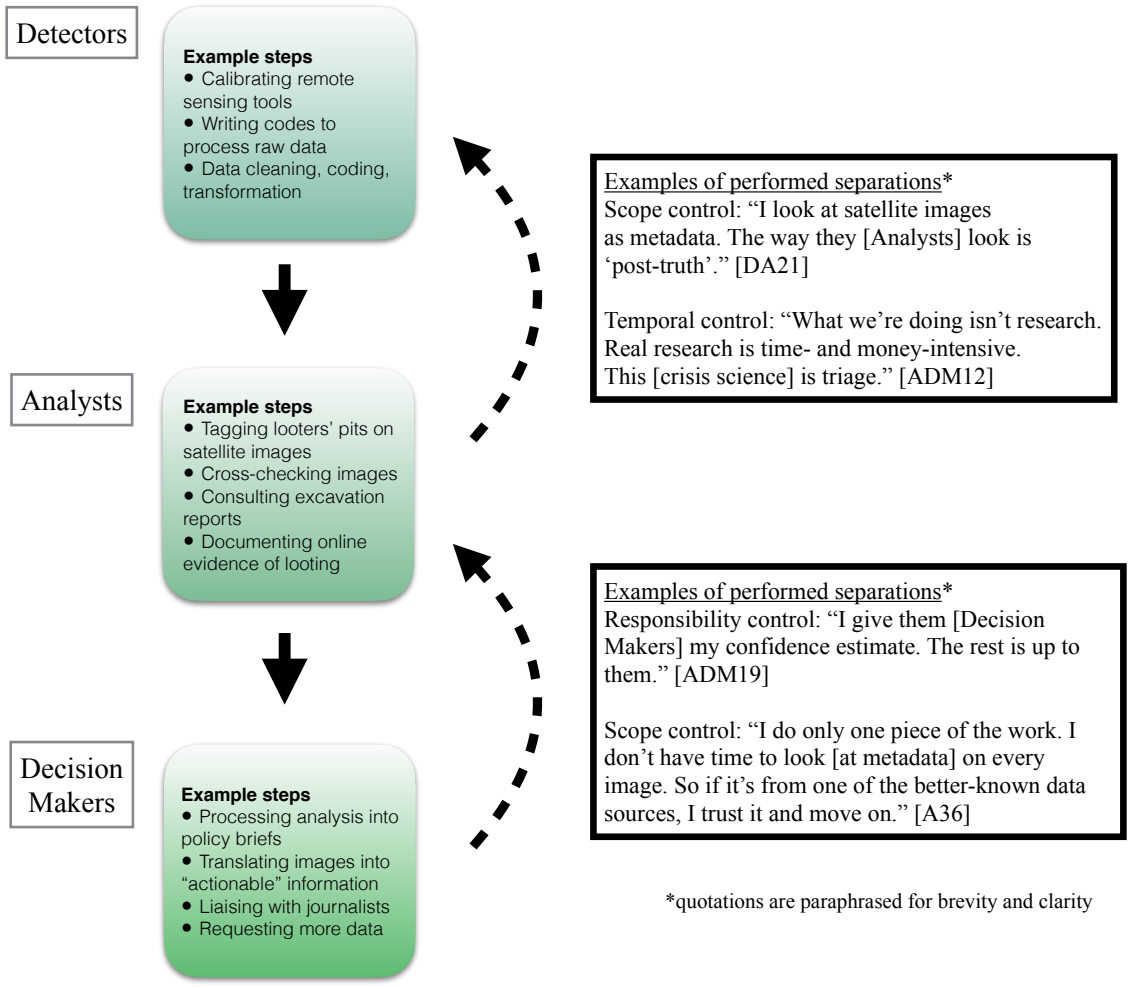

Figure S1. Production steps in conflict archaeology. Dashed arrows represent responses to crisis science reliability concerns.

Coding definitions

Figure 1 in the main text plots our respondents on a grid structured by field position, work commitment, and confidence measures. Near/Far and Embedded/Adjacent were binary coding decisions. As a result, each respondent can occupy only one quadrant. The spatial variation within each quadrant corresponds to the person's degree of confidence in the reliability of the research being produced by the crisis science response. For example, A32 was coded *Embedded/Near* because they performed crisis science analysis in close conjunction with Decision Makers ("Near"), and although they continued to perform non-crisis science research during the period in question, that other research was directly related to conflict archaeology. As such, their field position was "Embedded." Because A32 expressed moderate confidence in the overall research being produced by the crisis science response, they are plotted between High and Low Reliability (near the upper-right corner) of the *Embedded/Near* quadrant. To take another example, D27 was coded *Adjacent/Far* because they performed computational and programming tasks that were several steps removed from crisis science translational work ("Far"), and because those tasks were distributed across research projects including non-conflict archaeology projects ("Adjacent"). They expressed high confidence in the reliability of their work, and in the reliability checks performed by researchers closer to crisis science translational work. For this reason, they are plotted in the *Adjacent/Far* quadrant, in the High Reliability space. These and related coding decisions are explained at length in our data dictionary and coding analysis notes (available from the authors upon request).

Coding procedures

We performed two rounds of coding, with an inter-rater reliability (IRR) test between rounds. We used the Dedoose package for transcript storage and coding, structured by macro- and granular-level coding nodes. Two student research assistants transcribed the interview recordings, which were then checked by the authors for accuracy including technical terms and foreign language transliterations. All transcripts were de-identified and assigned a unique identifier. Macro-level nodes were established by the authors prior to Round 1 Coding, based on prominent themes in the interview schedule. Each author then coded 50% of the transcripts (randomly assigned). They first read a given interview in its entirety, then applied macro-codes, and created granular codes, during the second pass. The list of coding nodes is available from the authors upon request.

Once all of Round 1 coding was completed, the authors performed the IRR test on a 10% sample of excerpts. A stratified random sampling method with replacement was employed so that each Code received at least 2 excerpts selected. Replacement was allowed for efficiency, meaning that the same excerpt could be selected multiple times for each code. There were 35 unique codes and 1050 excerpts. 10% was 105 excerpts, which translated to exactly 3 rounds of sampling. (To see the specific excerpts attached to each ID, we created a "Round 1 Coding Excerpt Randomization" spreadsheet. This is also available upon request from the authors.)

The coding goal was not 100% agreement but rather to have sufficient overlap in coding to indicate an alignment in how the codes were being used and to ensure coding complementarity. Some divergence was expected and desirable as the coders have different backgrounds and interview experiences informing how they interpret the data. To ensure that both coders were in alignment, both coders reviewed patterns of code application with respect to three specific areas of interest and then compared them. The three areas of interest were: (1) scientific urgency, (2) structure of the conflict archaeology field, and (3) credibility and reliability in data/knowledge production. Based on the qualitative comparison, some adjustments were made to the data dictionary and coding decision rules prior to Round 2.

Round 2 coding was designed to be complementary coding rather than a full blind coding. This step relied on the coders having an IRR threshold score of at least 50%. Rather than re-reading and re-coding every interview, the coders reviewed each other's coding decisions, adding additional codes where needed and making comments where coding ideas diverged. We reviewed the Round 2 changes and reached agreement prior to analysis.

IRR equations and output are available upon request.