Beyond the existence proof: ontological conditions, epistemological implications, and in-depth interview research

Samuel R. Lucas

Abstract   In-depth interviewing is a promising method. Alas, traditional in-depth inter-view sample designs prohibit generalizing. Yet, after acknowledging this limitation, in-depth interview studies generalize anyway. Generalization appears unavoidable; thus, sample design must be grounded in plausible ontological and epistemological assumptions that enable generalization. Many in-depth interviewers reject such designs. The paper demonstrates that traditional sampling for in-depth interview studies is indefensible given plausible ontolog-ical conditions, and engages the epistemological claims that purportedly justify traditional sampling. The paper finds that the promise of in-depth interviewing will go unrealized unless interviewers adopt ontologically plausible sample designs. Otherwise, in-depth interviewing can only provide existence proofs, at best.

Keywords   Ontology · Epistemology · In-depth interviewing · Sampling · Probability sampling · Non-probability sampling · Snowball sampling · Purposive sampling · Theoretical sampling

What can we learn from in-depth interviewing? Surely, an interviewer can ask multiple respondents myriad questions, but what justifies attending to the answers or analysis? This question motivates the investigation. Because methodological justifications depend crucially on the match between ontological conditions and epistemological assumptions, addressing this question requires assessing the plausibility and coherence of key ontological and episte-mological assumptions of in-depth interviewing.

Although multiple assumptions ground every method, the study focuses on case-selection for in-depth interview (IDI) studies. Because all researchers must select cases for study, case selection is of broad interest.

S. R. Lucas (✉)
Department of Sociology, University of California-Berkeley, 410 Barrows Hall #1980, Berkeley CA, 94720-1980, USA
e-mail: Lucas@demog.berkeley.edu

To proceed, I first differentiate social science in-depth interviewing from other interviewing. Focusing thereafter on social science in-depth interviewing, I next convey a common feature of IDI studies. Then a fundamental ontological condition that affects all respondent selection strategies is offered. Afterwards, three case selection strategies are conveyed and interrogated. After addressing key myths of case selection, I consider two justifications of common IDI approaches—one claiming different generalization logics, the other claiming unique research aims, justify traditional strategies. The penultimate section addresses implementation. Summary reflections close the analysis.

1 Social science in-depth interviewing

1.1 Social science interviewing versus non-social science interviewing

Doctors interview patients, constables interview witnesses, reporters interview sources, and social scientists interview respondents. Doctors, constables, and reporters interview to learn what happened in a case of intrinsic interest to some client, e.g., a patient, victim, or citizen. If they investigate cause they seek causes of a specific incident, such as an illness, murder, or bribe; they do not aim to generalize to a class of events, although they may use others' generalizations (e.g., serial killers have an identifiable profile) in their investigation.

In contrast, social science interviewers look beyond "what happened?" Whereas a reporter or detective may ask whether wealthy person A bribed politician B, social scientists ask whether the wealthy unduly influence political decisions. From the social science perspective the bribed politician is perhaps one example, of so little specific interest that social scientists' reports can suppress the politician's name. For the reporter or detective, however, the politician's name is central to the story or criminal case.

Thus, social scientists' respondents are not necessarily intrinsically interesting or interested (Gomm et al. 2000, p. 102). And, owing to a focus on patterns about which respondents may be unaware, social scientists avoid directly asking respondents their research question (e.g., do wealthy people unduly influence your decisions?); instead, they probe unobtrusively.

1.2 In-depth interviewing versus survey interviewing

In-depth interviewing differs from survey interviewing. Suchman and Jordan (1990), survey interview critics, show that interviews are interviewer/respondent co-productions. This implies that respondents' cognitive and affective access to information may depend upon the co-constructed context. Consequently, different interviewers may elicit different answers from the same respondent, a possibility that can destabilize confidence in interview results.

Survey researchers address this challenge by standardizing training, questions, and responses to respondent queries. Yet, many factors reduce such efforts' success (e.g., Anderson et al. 1988). In-depth interviewers respond by tailoring questions to each respondent. Hence, IDI epistemology differs from that of survey research.

I focus on social science IDI and attend not to the interview but, instead, to the selection of interviewees. To that end, I first relate the common IDI approach to respondent selection.

2 The common strategy of in-depth interviewing: described and assessed

2.1 The common strategy of in-depth interviewing

Most IDI studies include a statement that serves as introduction to, justification of, and apology for the research design. For example, O'Brien (2008) writes:

> …to fully investigate the color-blind ideology that facilitates the process of whitening, however, and the color-conscious ideologies that would accompany the browning process, qualitative methods need to be brought to bear on the topic. Although such methods invariably yield a smaller sample, and one that cannot necessarily be generalized to a wider population, they ask key questions that large-scale surveys cannot. (p. 17).

Or, as another example, Orrange (2003) first writes:

> These data come from an in-depth study of forty-three advanced professional school students in law and MBA programs at a major research university in the Southwestern United States…. All respondents were recruited using snowball sampling techniques. (p. 456),

a design justified by the claim that:

> This study provides a rich contextual basis for an exploration into how these young adults think about meaning in life…. The in-depth interviews help us to evaluate whether their thoughts about the future, in general, and meaning in life, in particular, are merely whims and fancies or, on the other hand, are richly woven into the vocabularies of motive through which they discuss their aspirations. (p. 457).

However, Orrange qualifies every claim with the observation that:

> …some caution should be used in interpreting these findings, as the research sample is not necessarily highly representative of some well-defined population in the broader society; the analysis was based on research involving a purposive sample (p. 458).

This pattern of explicit description/justification of a design accompanied by admission that one cannot generalize the findings is endemic to IDI studies. Yet, after articulation, the qualification is apparently forgotten, for what invariably follows is a series of generalizations. For example, Orrange maintains that:

> These findings help us in clarifying the current state of affairs with respect to debates regarding the new individualism in both the narrow and the broad sense, while they also have implications for the postmodern challenge to the self. (p. 473),

and

> Certainly, one emerging and gendered alternative or adaptation to the potential tensions and conflicts some of the women face in forging careers and families is the "friends as family" alternative described…. This pattern of response represents one alternative to the dominant findings. (p. 474).

Sample design rendered respondents idiosyncratic subjects only, prohibiting using the findings to address general claims (e.g., *the* postmodern challenge to the self). Yet, Orrange suggests generalizing from them to larger societal patterns, abandoning the aforementioned caution.

Orrange's (2003) second generalizing excerpt counterposes allegedly dominant and alternative family patterns. Yet, as I show below, there is no way to summarize *within* the sample to determine a dominant pattern given the design.

As another example, O'Brien contends that:

> While qualitative work such as the present study may not be able to provide generalizable findings, it can suggest ways to better tailor surveys on this topic in the future. For instance, questions that ask respondents' racial and ethnic categories could be presented as a continuum (perhaps a scale of one to ten) that allows them to define for themselves how closely they identify with any particular group. (pp. 59–60).

O'Brien claims an inability to generalize, but on the basis of the analysis advises census takers and survey researchers how to pose questions to millions of future respondents, as if this advice does not constitute a generalization from the reports and experiences of O'Brien's respondents to future strategies for ascertaining others' experiences.

Although generalizing after admitting one cannot justifiably generalize is not the only challenge facing in-depth interviewing, it is perhaps the most visible sign that something is awry with IDI as commonly practiced.

2.2 Assessing the common in-depth interview strategy

I contend that in-depth interviewers generalize because realizing the social science value of IDI requires generalization. Caught in contradiction between sample design limitations and social science imperatives, interviewers, perhaps inadvertently, deploy synonyms for generalizing. Alas, a large supply of synonymic constructions is available.

For example, O'Brien (2008, p. 18) aims to "identify the particular experiences that are unique" to Latino/as and Asians as "middle" racial categories. Identifying unique experiences of a racial/ethnic group involves generalizing from sampled to nonsampled coethnic peers. How, after all, can one identify "unique Asian experience" without generalizing? Thus, the study aim could not be realized without generalizing.

Other synonyms include reference to more likely occurrences (e.g., Brunson 2010, p. 231), policy implications (e.g., Brunson 2010, p. 232), and theoretical implications (Wrigley 2002, pp. 48–49). To predict future occurrences is to offer conclusions for cases/instances one has not observed, i.e., to generalize, and policy- and theory-construction are generalizing activities by definition. These and hundreds of other published examples suggest there is no escape from generalizing; social science in-depth interviewers, like all other social scientists, are engaged in a process that requires one to generalize.[1]

Social scientists generalize *from* the sample and *within* the sample (Gomm et al. 2000, pp. 108–111). Unless they have intrinsically interesting cases (e.g., the 9 US Supreme Court justices), which might obviate out-of-sample generalization for some questions, social scientists generalize in both ways.

Accordingly, one may ask, of what value is a method practitioners admit prohibits generalization? Absent generalization, may IDI studies only help us understand the specific individuals interviewed? Most readers have no interest in O'Brien's, Orrange's, or most other respondents specifically. Thus, some way to generalize seems imperative.

As practiced, IDI studies actually have some limited value. Orrange noted some respondents referenced friends as fictive family, and O'Brien found some respondents both adopted

---

[1] Lincoln and Guba (1979,2000) claim generalization requires one to assume determinism. As Gomm et al. (2000, pp. 100–101) show, this claim is incorrect.

and abhorred racial prejudice. To the extent other study aspects are solid, both provide existence proofs; we learn that at least one person in the world held the described constellation of positions. But, given that human imagination permits infinite arrangements, is the ability to show that someone somewhere holds some position sufficient justification for frequent use of a method? If not, can IDI be arranged to provide more than existence proofs?

Because in-depth interviewing is difficult, interviewers can easily fail to produce even an existence proof. Still, I believe interviewers can go beyond existence proofs to generalization, a key possibility because fundamental aspects of the social render non-generalizing social science virtually impossible (Gomm et al. 2000, pp. 98–102). To generalize interviewers must adopt a coherent basis for generalizing, and resist claims that evade the inherent challenges of social research. The quintessential example of (likely unwitting) evasion is the claim that quantitative research generalizes to populations but qualitative research generalizes theoretically (e.g., Oberfield 2010), a claim that indicates the ontological impediments to and epistemological foundations of systematic knowledge production are insufficiently appreciated.

The tough challenges of research stem from the nature of reality. Although the nature of reality is contested, the feature of reality that concerns us is virtually uncontested. Thus, before considering the available design strategies, we need first delineate a key ontological condition all successful respondent selection strategies must address.

3 A fundamental ontological condition: social world lumpiness

The social world is ubiquitous; consequently, there is no asocial space within which social scientists can work or upon which they may stand and thereby avoid the inferential challenges inherent in the social world.[2] These unavoidable challenges are posed by, at minimum, the lumpiness and complexity of the social world. Strategies exist for addressing both features. Lumpiness is the primary culprit for case selection, so I focus on lumpiness.

All analysts confront a social world that is lumpy. By "lumpy" I mean that in the large-dimensioned social space there are concentrations of entities, and sparse locales; some constellations of characteristics are common, others rare; hills and mountains rise from some spots on the social terrain, valleys and ravines mark others. As Ragin (2008) rightly indicates:

> Naturally occurring social phenomena are profoundly limited in their diversity…. It is no accident that social hierarchies such as occupational prestige, education, and income coincide, just as it is not happenstance that high scores on virtually all nation-level indicators of wealth and well-being are clustered in the advanced industrial countries. Social diversity is limited not only by overlapping inequities of wealth and power but also by history. For example, colonization of almost all of South and Central America by Spain and Portugal is a historical and cultural "given" … (p. 147).

Though hundreds of permutations of conditions are logically possible, lumpiness drastically reduces the realized number. The feature also bedevils statistical research. Quantitative researchers' covariates are correlated, facilitating inadvertent generalization "off the support" (Neter et al. 1989, p. 262). For example, although some may earn $300,000, and some may

---

[2] The metaphorical social space may have far more than 3, 4, or 11 dimensions. Indeed, identifying and mapping its possibly changing dimensions is an on-going task of social science.

have 6 years of schooling, the combination of characteristics may not exist, problematizing all predictions for such logically possible cases.

Even as social world lumpiness justifies interest in existence proofs, it obscures both the living beings scattered in clusters across the social terrain and the social forces that cluster, scatter, enable, and constrain them. *If an analyst seeks to either count entities or apprehend (or interrogate) social relationships, social world lumpiness will prevent their success unless the study design addresses the threat social world lumpiness poses.*

Interviewers use respondents as entry points into or vantage points on this social world, seeking to discern respondents' statuses (e.g., predicaments, interpretations, trajectories, contradictions) and/or the social forces that partly pattern and produce these statuses. However, owing to social world lumpiness, observation of and from any vantage point (i.e., any respondent) will allow discernment "around" that point on the multidimensioned social terrain only. By implication, there is *no* vantage point of unassailably penetrating knowledge; the view from every location is partial, its horizon blocked by some feature(s) emerging out of or into the social plain. Consequently, every location is unable to reveal some aspects or force of the social world that would have been visible from elsewhere.

Further, a heretofore partially inscrutable mix of stochasticity and systematicity shape the social world and determine each specific location *and* the social forces and patterns visible therefrom. That the social world is partly determined by systematicity, and we incompletely understand it, means that if we (purposely or mistakenly) use any systematic social factor to select respondents, such as, for example, their membership in an accessible network, we likely bring a distorted set of social world patterns and force into view, because cases will be partly selected by social world lumpiness itself, which alone can distort our vision of the social world.

IDI studies usually report distributions of sample characteristics (e.g., two-thirds of the sample are women, 40% live in cities), but this information is insufficient. Because we lack knowledge of both the infinite factors underlying social world lumpiness and the connection between those factors and characteristics of sample members, it will remain impossible to include what would have been visible from relevant vantage points beyond our reach, such that we can neither assess nor eliminate the possibility that the inaccessible vantage points might be such that the power of the social forces they would reveal were they accessible would completely overturn our findings. Hence, our view of the social world may be horribly distorted, even as all to which we have access may appear sharp, coherent, "reasonable."

This result is not a matter of mathematics; it follows, instead, from social world lumpiness and our incomplete understanding of it. Thus, all research must address social world lumpiness.

4 Responses to lumpiness

Respondent selection is the key site where social world lumpiness is addressed. Setting aside experiments momentarily, in-depth interviewers may adopt one of three strategies: select everyone (i.e., census-taking), select respondents using a process independent of social world lumpiness (probability sampling), or recruit respondents in ways rife with social world lumpiness (non-probability sampling).

4.1 Census-taking

One response to social world lumpiness is to study the entire population, in which case anything one learns about the data perfectly generalizes to that population. Yet, censuses do not resolve the out-of-sample generalization challenge. To generalize to the population *between* censuses requires justification. Few analysts study a census because they are interested in that moment specifically. Different times can be different; thus, "I have a census" does not, by itself, justify generalization from census data. A basis for generalizing from censuses in such situations is described below (see Sect. 6.3).

4.2 Probability sampling: principles and implementation

Before probability sampling was developed many believed censuses were the only way to study social phenomena (Kruskal and Mosteller 1980; Desrosières 2004, pp. 210–235). The breakthrough of probability sampling allowed generalizing from study of comparatively few entities, if they are selected using probability sampling principles.

Probability sampling principles assume social world lumpiness, liberating research findings from its threat by giving every member of the *target population* a *non-zero* and *knowable* probability of selection into the sample. Assigned probabilities are independent of any causes of social world lumpiness not explicitly part of the probability assignment process. Sampling using those probabilities produces samples that collectively have no distorting relationship to social world lumpiness.

Identifying the target population—the set of entities to which the sample most directly generalizes—is not objective. Researchers must use their judgment, and reasonable analysts may disagree. For example, to study U.S. husbands some analysts might treat all males in same-sex marriages as husbands, others might exclude them, others might take one husband per male-male marriage, others might allow the couple or individual to decide, and there may be other options, too (e.g., including women in same-sex marriages who view themselves as husbands, whatever "being a husband" may mean). Operational definitions are not objective; still, an explicit, systematically-defined target population is necessary. One might chafe at this requirement, but inferences require some scope for their first-order applicability.

Because target population members' selection probability must be non-zero, it must be possible to determine or estimate every target population members' inclusion probability, else sampling statisticians could not confirm satisfaction of the non-zero probability requirement. All else equal, the more target population members with zero chance of inclusion, the more biased the sample.

Although all selection probabilities must exceed zero, selection probabilities may differ. Indeed, unequal selection probabilities are common. Often analysts want to increase representation of rare categories. For example, if a 20 husband sample contained only 1–2 husbands in interracial marriages, within-sample comparisons would be imprecise. If comparing husbands in same-race and interracial marriages were a research aim, analysts might oversample husbands in interracial marriages to facilitate within-sample comparisons, perhaps sampling 10 husbands of each type. This oversampling roughly equates the precision of inferences about each kind of husband. However, to generalize to all US husbands one must downweight the responses of husbands of interracial marriages, else findings will be based on an assumption that 50% of US marriages are interracial. Yet, given a sample with knowable non-zero probabilities of inclusion for target population members, one may easily weigh responses differently when estimating overall incidences versus when comparing some within-sample groups.

4.3 Non-probability sampling

An alternative, non-probability sampling, does not require identification of a target population, members of what would have been the target population may have zero chance of inclusion, and it may be impossible to calculate inclusion probabilities. It is well-known that such samples prohibit out-of-sample generalization. What is less commonly known is that such samples prohibit generalization *within* the sample, i.e., no comparison within or summary of the sample (e.g., identifying the dominant experience) is justifiable.

Non-probability sampling necessitates adopting implausible ontological or epistemological assumptions. Assuming social world homogeneity—a smooth as opposed to a lumpy social world—would justify non-probability sampling. In a smooth social world every vantage point allows unhindered access to all social forces and patterns. The social world homogeneity assumption seems most coherent if all persons are understood to behave, react, feel, respond *exactly* the same such that the connections amongst phenomena are unvarying.

Regardless, social world homogeneity is implausible. Indeed, many draw inspiration from those who first rejected this assumption (e.g., Boas 1896, especially 904–905) and look back in sadness or even horror at those who journeyed to foreign shores brazenly clutching their homogeneity assumption (of, for example, a universal sequence of societal development). Using an ontological assumption of social world homogeneity to justify non-probability sampling is to commit the same error.

Another way to justify non-probability sampling is to assert its invulnerability to social world lumpiness. Examples below document, however, that this assertion is false.

Snowball sampling. Snowball samplers recruit and interview some volunteers, afterward asking for referrals to other potential respondents, perhaps even requesting referrals to persons with specific characteristics. In this way the sample may cumulate, like a snowball.[3]

Berg (2006) notes that snowball sampling gives more socially connected people higher selection chances and assigns some target population members zero inclusion probability because they do not share networks with the type of people from which the snowball starts. Further, all inclusion chances are unknowable because unsampled persons' network characteristics are unknowable. These factors undermine all generalization.

IDI analysts acknowledge the prohibition on out-of-sample generalization, but snowball samples also prohibit within-sample summarization. To learn what is dominant within the sample it must be defensible to count or cumulate respondents' responses. In order to sum(marize) respondents' responses all respondents must have a knowable probability of inclusion, because the reciprocal of this probability determines the weight each respondent's responses should receive.

For example, in the imagined probability study of husbands above, assume the population has 900 same-race couple husbands (SRCHs) and 100 interracial couple husbands (IRCHs). Because of IRCH oversampling, each SRCH has a $10/900 = 1/90$ chance of sample inclusion; each IRCH has a $10/100 = 1/10$ chance of inclusion. Thus, each sampled SRCH represents 90 others and each sampled IRCH represents only 10 others. When comparing these groups we may ignore this difference. However, to determine "What common orientations to marriage do husbands strike?", à la O'Brien's analysis of the racial middle, we need weigh each SRCH response as if it was reported 90 times and each IRCH response as if it was ascertained 10

---

[3] Respondent-driven sampling, a proposal to repair snowball sampling (Salganik and Heckathorn 2004), requires larger samples than usual for IDI, making the solution's feasibility for IDI unclear. Further, questions about the success of the method remain (e.g., Goel and Salganik 2010). Also, snowball samples are not how one should sample networks. Instead, probability sample persons (or nodes) and extend outward.

Table 1 Hypothetical interviewees, marriage orientations, and typical and selected possibly appropriate weights

Respondent	Orientations	Typical weights	Selected possibly appropriate weights		
Column # 1	# 2	# 3	# 4	# 5	# 6
1	■	$100 \Rightarrow 1$	52	150	180
2	□ ■	$100 \Rightarrow 1$	52	150	180
3	○ ❖ •	$100 \Rightarrow 1$	250	150	180
4	○ ▣	$100 \Rightarrow 1$	14	150	20
5	□	$100 \Rightarrow 1$	52	150	180
6	• ○ ■	$100 \Rightarrow 1$	250	50	180
7	▣	$100 \Rightarrow 1$	14	50	20
8	○ ▣	$100 \Rightarrow 1$	52	50	20
9	□ • ❖ ▣	$100 \Rightarrow 1$	250	50	20
10	▣ □ △ •	$100 \Rightarrow 1$	14	50	20
Dominant Orientation		▣	•	○ & □	■
Sum of the Weights		1000[a]	1000	1000	1000

[a] With equal weights, the implied ($\Rightarrow$) effective "comparative" weight is 1

times. Otherwise we may misidentify the dominant orientation. The example is simple, but all complexities (e.g., weighting for nonresponse) reach the same conclusion.

One must know selection probabilities in order to produce many within-sample claims, from exact numerical counts (e.g., 74%) to rough approximations (e.g., the dominant view). Indeed, one even must know the selection probabilities to determine whether one needs to know the selection probabilities. Absent this knowledge one cannot justify *any* weighting, including equal weighting, of respondents' responses.

Table 1 illustrates these principles' consequences. In 10 interviews seven distinct marital orientations are expressed, symbolized by •, ○, ■, □, ▣, ❖, and △. Column 3 contains typical, equal, weights, a pattern that requires all target population members have equal selection chances. Because snowball sampling violates this criterion (Berg 2006), we know equal weighting is incorrect, even though most in-depth interviewers implicitly use equal weighting.

In Table 1 the dominant orientation under the typical weighting is ▣ . Yet, if the correct weights are in column 4, then the dominant orientation is •. Other weightings produce other dominant orientations.[4] Findings are sensitive to weighting, yet—and this is the problem—non-probability sampling provides no basis for selecting any one of the infinite number of possible weighting patterns. Thus, the validity of any summary of respondents' orientations cannot be established. All one can validly do is state these orientations exist. One cannot even conclude that unobserved orientations are uncommon for, with non-probability designs, absence offers absolutely no evidence of incidence. Thus, when Orrange (2003) claims to find *dominant* and *alternative* perspectives, neither designation is defensible.

It may seem counterintuitive that one needs information dependent upon the target population—the selection probabilities—to make *within*-sample claims. But, this necessity is not really odd. If one taste-tested several liquids to evaluate an establishment, one's evaluation

---

[4] I treat each orientation separately; using combinations would also show weights matter.

would depend, in part, on knowing whether a liquid was coffee, tea, beer, chardonnay, viognier, or zinfandel. Two conclusions follow. First, at the beverage level, a good coffee would be a terrible viognier—understanding a case depends on knowing what the case represents. Second, the summary would depend on what each sampled liquid represented and how much each beverage "represented" the establishment. A bad viognier might not lower our evaluation of a coffeehouse much, but it might lower our evaluation of a wine bar significantly. Hence, to understand each sampled case and to summarize the full sample one must know each unit's representativeness. Selection probabilities, only available for probability samples, provide that information.

Other Non-Probability Sample Designs. Such problems hound all non-probability designs, including two other IDI mainstays—purposive sampling and "theoretical" sampling. For Marshall (1996), in purposive sampling the "researcher actively selects the most productive sample to answer the research question (p. 523)," picking and choosing respondents based on their view of who will aid their research. This might require the researcher to develop:

> a framework of the variables that might influence an individual's contribution…based on the researcher's practical knowledge of the research area, the available literature, and evidence from the study itself. This is a more intellectual strategy than the simple demographic stratification of epidemiological studies, though age, gender, and social class might be important variables…. It may be advantageous to study a broad range of subjects (maximum variation sample), outliers (deviant sample), subjects who have specific experiences (critical case sample) or subjects with special expertise (key informant sample)…. During interpretation of the data it is important to consider subjects who support emerging explanations and, perhaps more importantly, subjects who disagree (confirming and disconfirming samples) (p. 523).

In theoretical sampling, a kind of purposive sampling, interviewers build theories during data collection, selecting new respondents in order to interrogate emerging theoretical positions.

Purposive sampling sounds valuable. After all, who can oppose drawing a sample that will allow the researcher to answer their question? At issue, however, is what procedure will most likely satisfy this aim. The problem with purposive sampling is that social world lumpiness and its basis in infinite factors of unknown power mean that no matter how much information interviewers have, distortion is likely because they still lack sufficient knowledge and power (i.e., omniscience and omnipotence) to render impotent the infinite, unknown determinants of social world lumpiness. Thus, purposive sampling almost assures they will *not* be able to validly answer their research questions, even as the design may instill undue confidence in the unknowably distorted picture the research produces.

Yet, for example, there is a difference between asking what people feel about car disrepair and asking what *mechanics* feel about car disrepair. To generalize to mechanics' views one need not purposively sample; instead, define a target population (mechanics) and probability sample within it. This is a perhaps subtle but crucial distinction. With this target population one would be as mistaken to include non-mechanics as to exclude novice mechanics. Concerning the latter, expecting expert mechanics to know all mechanics' answers, or worse, giving novice mechanics' views zero weight, is unjustifiable. One cannot purposively select *which* target population members to interview, for doing so reinforces vulnerability to social world lumpiness.

All ostensibly sensible purposive sample designs have probability sampling analogs. For example, Marshall (1996, p. 523) and Small (2009, p. 13) advocate purposively sampling for range by recruiting subjects at variables' extreme values. Stratified sampling is the prob-

ability sampling analog; sampling at different rates depending on a stratifying variable's values provides the range sought *and* generalizabilty. Or, Small (2009, pp. 24–25), incorrectly claiming probability sampling requires setting sample size *a priori*, prefers sampling non-probabilistically until a conclusion emerges. Sequential sampling, a time-tested design (Wald 1945), is the probability sampling analog. Accordingly, analysts need not purposively sample to accomplish their purpose.

Basically, non-probability sampling rejects the zen of probability sampling—an openness to obtaining information from any target population member—in favor of a controlling, restrictive, possibly even arrogant[5] search for the "best" respondents, as if that designation itself is not driven by social world lumpiness, and as if the allegedly "less-than-best" have no information or experience analysts need access, consider, or respect.

5 Myths of probability sampling

Non-probability sampling is of extremely limited utility, providing grounds, at best, only for existence proofs. Yet, many interviewers continue to use such samples for more. Several myths about probability sampling seem to justify acceptance of non-probability sampling.

One myth is that small samples of any kind prohibit generalization (e.g., Marshall 1996, p. 523; Byrne 2001, p. 494; Small 2009, pp. 11–13), such that one may as well draw a non-probability sample. In reality, however, sample size has nothing to do with generalization. As has been known since the infamous "Dewey Defeats Truman" headline, generalizability is determined by the sampling process, not the sample size. Even a probability sample of 1 allows defensible, though perhaps imprecise, generalization; after all, one may generalize to a bottle or even a vintage with one sip. Further, while large sample size is one route to precision, there are other routes to this goal. For example, with effective probes in-depth interviewers can often attain *greater* precision than surveys might (Suchman and Jordan 1990). Such possibilities mean that small probability samples *can* provide precise evidence. Thus, in-depth interviewers need not apologize for small sample sizes—for precision, sample size is not necessarily determinative, and for generalization, size does not matter.

Another myth is that probability sample designs rarely produce sufficient numbers of uncommon groups (e.g., Small 2009, p. 13); such claims imply probability sampling must offer everyone equal selection chances. As noted above, probability sampling selection chances can vary by persons' characteristics, perhaps to boost the sample incidence of rare populations.

Although many use non-probability sampling although a list is accessible (e.g., Orrange 2003), another myth is that unless one can list all population members one cannot draw a probability sample (e.g., Marshall 1996, p. 523; Montemurro and McClure 2005). This myth is not entirely false, but because a list of lists is itself a list of the primary elements, the opportunity for probability sampling is legion despite this seemingly demanding requirement. So, for example, O'Brien (2008) could have first sampled jurisdictions, then neighborhoods within those jurisdictions, residences within those neighborhoods, and then persons within those residences. Those failing to meet the study's racial inclusion criteria would be omitted at the last stage. By varying selection probabilities O'Brien could have oversampled neigh-

---

[5] Despite the obvious intransigence of social world lumpiness and its power to obscure social forces and relations the purposive sampler asserts, "I as interviewer know who can help me answer my research questions and who I can ignore," thus producing bias because the interviewer *is* biased against studying (would be) target population members the interviewer believes will not contribute to the research.

borhoods with more Asians and Latino/as to increase the design's logistical efficiency. Such multistage sampling can be straightforward, and allows one to generalize findings.

Another myth is that "for a true random sample to be selected, the characteristics under study of the whole population should be known (Marshall 1996, p. 523)." Actually, to draw a random sample of any entity one need only be able to identify the entity in the sampled context. Thus, one might think probability sampling gays and lesbians in a city is impossible. Yet, stratifying neighborhoods by an estimated density of gays and lesbians, and screening upon first contact, could secure a probability sample.[6]

Obviously, success with this strategy depends crucially on interviewer ability to secure cooperation. Fear of the slammed door may underlie some analysts' resistance to probability sampling, for with non-probability sampling one can approach acquaintances, volunteers, and other "approachable" subjects only. It should be obvious that such operations guarantee bias. Further, survey interviewers, facing the prospect of slammed doors for decades, have successfully secured subjects' cooperation for even sensitive topics (e.g., Das and Laumann 2010), and I doubt in-depth interviewers are less creative or socially adept. Regardless, securing cooperation is an unavoidable part of interviewing. Though the prospect of securing cooperation for probability sampled subjects may be intimidating, all research has its difficult moments; interviewer fear or discomfort offer no justification for non-probability sampling, a design whose comparative ease of accumulating interviews is irrelevant owing to its nearly wholesale analytic ineffectiveness.

An especially pernicious myth claims that probability sampling "is likely to produce a representative sample only if the research characteristics are normally distributed in the population (Marshall 1996, p. 523)." This is untrue. For example, votes in a two-candidate race are not normally distributed because dichotomous variables cannot be normally distributed (Kokoska and Nevison 1989, pp. 1, 6). But, probability sampling will produce samples that accurately estimate each candidate's vote proportion, thus effectively representing the voting population.

Finally, another myth is that probability samples are unnecessary unless one seeks to find the "average" experience (Wright and Copestake 2004, p. 360). This myth may originate in social science statistics classes that teach sampling using means. Using means, which students understand, should help students grasp the new information, the implications of sampling. With this pedagogically reasonable strategy, however, students may never realize that the conclusions apply to assessments of *any* and *every* characteristic of the social world—univariate distributions, trends, linear and nonlinear relationships between variables, patterns of clustering, and more.

To consider this issue I constructed a population of 1,000,000 cases with measures on four variables, setting their exact linear intercorrelations. The linear correlation coefficient, R, measures the relationship between two variables, and ranges from -1 to 1. Positive (negative) values mean that higher (lower) levels of X_j go with higher (higher) levels of X_k; zero signifies no linear relation. Statistical relations reflect substantive (e.g., in 2010 children of the wealthy were more likely to enter college than were children of the poor) and theoretical (e.g., cultural capital aids educational success) claims and thus are of broad interest.

Few in-depth interviewers aim to estimate R. Still, the exercise addresses two key issues. First, if nonprobability samples fail to capture relations between variables, it suggests one

---

[6] Novice analysts often severely narrow their study population, defining it via specific values of several variables of interest (e.g., education, race, age, sex, sexual orientation, number of children, and birth order). While not wrong, the approach complicates sampling, reduces study relevance, and undermines inferences about each dimension that defines their target population.

needs probability samples to assess relationships, however identified, not simply to estimate "averages."

Second, IDI studies often aim to interrogate the processes that underlie social relations. The implicit epistemological claim of non-probability sampling advocates is that non-probability samples will mirror those relations and processes. I contend, however, that because non-probability samples are unlikely to reflect social relations accurately, they are unlikely to reflect the processes beneath those relations, either. If I am correct, non-probability samples are unsatisfactory. So, for example, if in the population members of Group A tend to be hostile to immigrants, then, I contend, one should avoid methods that produce samples biased such that Group A members welcome immigrants, because in such samples the processes one observes are likely to be "off" as well, leading one to mischaracterize processes producing attitudes toward immigrants. Thus, even interviewers uninterested in estimating R should care whether a sample design is unlikely to match the real-world relations they seek to interrogate.

I drew 10 different probability samples of 40 cases each from the population of 1,000,000, obtaining the Rs for each. I then repeated the exercise for 10 different snowball samples of 40 cases each.[7] Table 2 contains the results obtained for each sample, with the actual population results arrayed across the top.

Probability sample estimates bracket the population values but the range is large, averaging .4047. With only 40 cases, a large range—low precision—is expected, just as some qualitative researchers rejecting probability sampling maintain (e.g., Small 2009).

Despite the imprecision, however, probability sample results match the population relations; 59 out of the 60 coefficients correctly estimate the direction of the correlation, an error rate of less than 2%. Thus, the probability samples, although small, reproduce the population relations in-depth interviewers seek to interrogate.

Half the non-probability sample Rs have larger ranges than their probability sample counterparts, suggesting that snowball sampling is not consistently more precise. More troubling, however, is that the non-probability samples are horribly biased, as high and low estimates never bracket the population value. In fact, the most precise estimates are extremely biased. The bias is consequential in that 29 of the 60 snowball sample estimates are wrongly signed (e.g., negative when in the population it is positive). Indeed, every snowball sample had multiple wrongly-signed coefficients, suggesting that IDI snowball sampling may commonly mis-estimate population relations (e.g., children of the wealthy are *less* likely to enter college than were children of the poor, cultural capital impedes educational success). It is unlikely that processes excavated through study of such samples would provide leads analysts should further interrogate. Indeed, using findings from such horribly biased samples to guide research could easily send analysts off on multiple, literally misleading investigations.[8]

Probability sampling succeeds because in a population of one million there are 9.99×10^{239} probability samples of 40, which is the number 999 followed by 237 zeros. Amidst this vast sea of samples any obtained sample is likely to be typical. In contrast, each of the incalculably fewer non-probability samples is necessarily atypical, for networks concretize social world lumpiness. The only snowball samples possible are those that networks can pro-

---

[7] The probability sample weights are equal by design. As correct weights for snowball samples are unknowable, to match standard practice I weight each such respondent equally. Further simulation details, data, and files are available at http://www.SamuelRoundfieldLucas.com/LucasQandQBEP.php.

[8] One might try salvaging non-probability sampling by focusing on the atypical, such as immigrant-welcoming Group A persons, but the findings imply the sample will not represent immigrant-welcoming Group A persons, either.

Table 2 Correlation coefficients, 4 different variables, repeated sampling

Population value →	.20 $X_1 X_2$	.45 $X_1 X_3$	.55 $X_1 X_4$	-.20 $X_2 X_3$	.60 $X_2 X_4$	.65 $X_3 X_4$	1,000,000 n
Probability samples							
1	.2771	.2853	.5065	−.2867	.6782	.5001	40
2	.5225	.1502	.5532	−.0954	.7104	.6205	40
3	.4723	.3310	.6428	−.1580	.6821	.6060	40
4	.0528	.4342	.5406	−.4301	.4163	.6259	40
5	.2909	.3162	.5624	−.2560	.5787	.6251	40
6	.0591	.7565	.6619	−.1482	.5889	.7046	40
7	.0587	.6347	.6175	−.2340	.5340	.6862	40
8	.2528	.5546	.6959	−.3368	.6547	.4832	40
9	.1453	.3043	.4416	−.2657	.5440	.6489	40
10	−.1178	.5752	.4926	−.1806	.4568	.7817	40
Range of estimates	.6403	.6063	.2543	.3347	.2941	.2985	$\Sigma\lvert$Parameter$_{jk}$ − Mean Est$_{jk}\rvert/6$
Mean of estimates	.2014	.4342	.5715	−.2392	.5844	.6282	.0192
Snowball samples							
1	−.1833	.0084	−.0692	−.9423	.0474	−.0543	40
2	−.1298	.0474	−.0871	−.9658	−.0512	.0057	40
3	−.0807	−.0097	.0741	−.9646	.2298	−.2376	40
4	−.1470	.0790	−.1191	−.9709	.0561	−.0023	40
5	−.3015	.2192	−.1597	−.9742	.1451	−.0955	40
6	−.2882	.2884	.2280	−.9727	.0236	−.0938	40
7	.0390	−.1077	−.0484	−.9601	−.0142	.0376	40
8	−.1891	.1352	−.0011	−.9732	.0594	−.0590	40
9	−.1206	.0913	−.1082	−.9697	−.2733	.2741	40
10	−.0452	.0300	−.2251	−.9842	.0751	−.0593	40
Range of estimates	.3405	.3961	.4531	.0419	.5031	.5117	$\Sigma\lvert$Parameter$_{jk}$ − MeanEst$_{jk}\rvert/6$
Mean of estimates	−.1446	.0782	−.0692	−.9697	.0474	−.0543	.5558

duce. Thus, snowball sampling processes are suffused with social world lumpiness-produced distortions.

Lastly, one way to increase the value of small sample-size studies is to combine the samples with others as they accumulate. Alas, each non-probability sample stands alone; one cannot justify combining them. Were one to combine them anyway results are still biased, as the Table 2 row of non-probability sample means indicates—all are far from the population parameter, and three are wrongly-signed. However, one can combine probability samples; when one does so here, all means are correctly signed, and all fall within .0392 of the population parameter. Thus, non-probability samples not only misallocate the primary analysts' time, but also they do not constitute an investment future scholars may exploit.

The main aim of the exercise was to address whether probability samples are useful only for estimating "averages." The answer is no. Probability samples are essential for study-

ing social relations. Thus, even if one is uninterested in calculating Rs, if one seeks to probe social relations, non-probability samples fail to reproduce the relations one seeks to study, sabotaging the effort to understand the processes/mechanisms behind social relations.[9]

6 Generalization logics[10]

Some analysts claim that certain generalization logics justify non-probability sampling. Can logics resolve the challenges social world lumpiness poses? Yin (2009) identifies 3 generalization logics: sample to population extrapolation (SPE), analytic generalization (AG), and case-to-case transfer (C2CT). We consider each as well as another proposal.

6.1 Sample to population extrapolation

Probability sampling and random assignment in experiments differ, but they have similarities. Considering those similarities will pay dividends.

Despite popular usage of the term "experiment" (e.g., the politician who admits "experimenting" with drugs in college), every exploration is not an experiment. Two definitive features of behavioral experiments are: (1) random assignment of research subjects to groups, and, (2) manipulation which exposes groups to different treatments (Campbell and Stanley 1963, pp. 6–42).

Experiments allow generalization because the experimenter *randomly* allocates entities to treatments, thereby canceling out distortions social world lumpiness produces. After treatment the experimenter obtains the difference in groups' outcomes, which the experimenter views as reflecting the treatment's causal power—a theoretical claim—and may posit as applicable outside the laboratory—an extrapolation to a population.

Consequently, repeating an experiment with slight adjustments (e.g., different subject pools (sophomores, managers)) maps the findings' extensiveness, just as probability sampling different populations (e.g., Californians, Czechs) can map the extensiveness of probability-sample based findings. Undistorted mapping is possible because both random assignment and probability sampling tame distortions produced by social world lumpiness. Henceforth, therefore, we can regard probability sampling, random allocation, and censuses (which assign everyone a known, nonzero (i.e., 100 %) selection probability) as *probability methods*.

6.2 Analytic generalization

Yin (2009) contends that AG logic is fundamentally distinct from SPE. Small (2009, pp. 24–27) agrees, and argues this "case study logic" applies to IDI. Our question does not concern AG in general but, instead, AG as an epistemological justification for non-probability sampling for IDI, as Small (2009) claims. Yin (2009) states:

> Critics typically state that single cases offer a poor basis for generalizing. However, such critics are implicitly contrasting the situation to survey research, in which a sample is intended to generalize to a larger universe. *This analogy to samples and universes is*

---

[9] One cannot salvage non-probability samples by confirming sample correlations are correctly-signed because Table 2 also implies unobserved correlations may be compromised.

[10] Arguably, logics of generalization precede case selection principles. Yet, confusion about the impact of social world lumpiness and how probability sampling addresses it required treating case selection first, for this discussion requires an accurate understanding of sampling.

incorrect when dealing with case studies. Survey research relies on *statistical* general-
ization, wheras [sic] case studies (as with experiments) rely on *analytic* generalization.
In analytic generalization, the investigator is striving to generalize a particular set of
results to some broader theory. (44, emphasis in original)

Seeing an SPE/AG distinction as justifying non-probability sampling implies probability
sampling renders units asocial, as if the act of probability sampling someone separates them
from the social processes and forces that animate their experience. In reality sampled persons
remain socially embedded, and collectively re-present to analysts the social world's complex
patterns and relations. Thus, one may draw valid inferences about the social forces that move
sampled persons, and extrapolate the substantive and theoretical findings to the population
and society.

The suggestion that AG is so fundamentally different from SPE that it justifies non-
probability sampling therefore collapses. When shorn of probability sampling, census-taking,
or random assignment, the use of AG for IDI falls prey to the problems already demonstrated
with non-probability sampling. Yin (2009), noting that generalization is not automatic, con-
tends that:

A theory must be tested by replicating the findings in a second or even a third [case]….
Once such direct replications have been made, the results might be accepted as pro-
viding strong support for the theory, even though further replications had not been
performed. This *replication logic* is the same that underlies the use of experiments
(and allows scientists to cumulate knowledge across experiments). (p. 44, emphasis in
original).

Although some embrace Yin's analogy with experiments (e.g., Small 2009, pp. 25–27), the
analogy is mistaken. As noted above, what makes replication for experiments work is random
allocation. Thus, replicating non-probability sample studies (or interviews) is to engage in
"cargo cult" ritualism (Small 2009; Feynman 1985, pp. 308–317), using a method's form
while failing to grasp and satisfy its essentials. Repeatedly using non-probability sampling
only replicates the same biases produced by the same inaccessibility of certain aspects of the
social world that hampered the original study owing to social world lumpiness.

This result was visible in Table 2; in at least seven of ten non-probability samples X_1 and
X_2, X_1 and X_4, and X_3 and X_4, are negatively related. However, the true relations are positive;
the replications certify falsehoods, indicating that replicating non-probability sample studies
is easily misleading.[11] To use AG logic interviewers must use probability methods.

6.3 Case-to-case transfer

A third generalization logic, C2CT, entails comparing a sample's characteristics with those
of a new case (Gomm et al. 2000, pp. 105–106). If the characteristics match, one generalizes
study findings to the new case. With C2CT one may extrapolate findings from census data
to a proximal year.

This is a solid generalization logic. However, to use it study findings must be solid. Because
we have already shown that Orrange has no basis for summarizing the non-probability sam-
ple (see Sect. 4.3), to extrapolate from Orrange (2003), for example, one must decide which
of the 43 idiosyncratic narratives to apply. This impediment to C2CT will haunt every non-
probability sample IDI study. Yet, if one uses probability methods C2CT differs little from

---

[11] Varying the replication also does not work.

SPE. Both entail generalizing from studied to unstudied cases. One must always justify such extrapolation.

Even quantitative studies do not probability sample for every level; few researchers randomly select a country for study, for example. One cannot use SPE logic to generalize research on US residents to other countries if the US case was selected non-probabilistically, even if US residents were probability sampled. To extrapolate findings to another nation one may use C2CT logic; to do so one must establish that the nations are similar enough to justify extrapolation. Such assessments require researcher judgment. Yet, judgment is useful only if the original study produced solid summaries and evidence, and only probability methods do so.

6.4 Natural science logic

One might contend that natural scientists do not use random assignment in their experiments; thus random assignment is not definitive of experiments; consequently, an analogue—probability sampling—is unnecessary for non-experimental research.

This view forgets that both natural and social scientists encounter legitimation challenges in a context of contestation. The trappings of scientific legitimacy are limited and include a language of experimentation. Thus, for example, astronomers claiming to experiment with rather than observe the sun (e.g., Price 1995) may be positioning their research for maximum legitimacy, making the language a legitimation strategy, not necessarily a coherent epistemological position.

However, where it is a coherent epistemological position, the ontological basis for natural scientists' rejection of probability methods conflicts with social science ontologies. Natural scientists presume their material is homogenous; for example, they assume all neutrinos are the same. Assuming away heterogeneity leaves natural scientists with no distorting lumpiness to address. To use this natural science epistemology social scientists must make the same ontological move, which would entail assuming away the heterogeniety—*social* world lumpiness—that partly justifies distinguishing social and natural science. Indeed, accepting this natural science logic for social analyses seriously weakens the ground for interpretive social science.

One could affirm both physical world homogeneity and social world heterogeneity. For example, in non-experimental early drug development studies (1) interchangeable rats (2) substitute for humans; both moves assume physical world homogeneity. In later drug trials, however, natural scientists randomly assign treatments, thereby addressing (social world) heterogeneity. These superficially contradictory research operations are reconciled if natural scientists assert physical world homogeneity, affirm social world lumpiness, and deny a social realm amongst non-human animals. Of course, this means that natural scientists accept the necessity of probability methods in social research.

6.5 Generalization logics

One could construct countless defenses of non-probability sampling by extracting a claim from an epistemology within which it coheres to insert into a domain constituted by incompatible ontological conditions,[12] an error similar to that some statistical researchers commit when they deploy quantitative research practice to dismiss *all* qualitative research. The similar structure of the approaches reflects that there is no end to the claims that, standing alone,

---

[12] For example, one could claim non-probability sampling works for Mill's methods; this claim falsely analogizes comparative historical and IDI research.

could appear to justify a method. But research is complex—no one claim, standing alone, can justify anything.

Accordingly, analysts must forge chains of reasoning to craft coherent epistemologies well-matched to ontological conditions. Proceeding otherwise can create ontology-epistemology mismatches that, ultimately, may undermine social science or, at least, interpretive social science. Defending non-probability sampling for IDI in this way will provide, at best, a pyrrhic victory.

7 Different research aims

A final way to justify non-probability sampling is to contend that IDI research questions differ fundamentally from those for which probability sampling is used. Upon closer inspection, however, these claims are unsustainable.

7.1 Interpretation of reality

Some interviewers seek to interpret reality, where reality is seen as "constructed, multi-dimensional, ever-changing; [such that] there is no such thing as a single, immutable reality waiting to be observed and measured (Merriam 1995, p. 54)." Interestingly, quantitative analyses of probability samples often reveal persons' interpretations and prospects systematically and consequentially differ (e.g., Schuman and Rieger 1992; Lucas 2008, pp. 23–52), suggesting that constructivist ontology and probability sampling are compatible.

Indeed, a constructivist must resolve whether the multiple interpretations they excavate evidence a multiplex social world or, instead, appear only because social world lumpiness has distorted the researcher's data. Social world lumpiness offers a competing explanation for constructivists' findings, reinterpreting them as error. Thus, to solidify findings of multiplicity, methods that address rather than ignore social world lumpiness seem imperative. Consequently, even if the aim of interpreting reality differs from the aims other analysts pursue, probability methods seem essential.

7.2 Understanding

Another claim is that in qualitative research "improved understanding of complex human issues is more important than generalizability of results (Marshall 1996, p. 524)," that "studying a random sample provides the best opportunity to generalize the results to the population but it is not the most effective way of developing an understanding of complex issues relating to human behavior (Marshall 1996, p. 523)."

The question to pose to these assertions is: How can one understand a complex human issue without generalizing from finite observations to a broader human condition? Seen in this way, "understanding of complex human issues" is just another phrase for "generalization about complex human issues." If we cannot distinguish between these two phrases we must reject the supposition that researchers who seek the former are liberated from the constraints entailed in generating the latter.

7.3 Different research aims revisited

This final aim returns us to our beginnings, the ineluctability of generalizing in social science. Once we realize generalization is unavoidable, we must respond to the threat social world lumpiness poses. Claiming to possess different aims is an unsuccessful response.

Rejection of non-probability sampling raises a final question—how may interviewers probability sample for IDI?

8 Implementing probability sampling for in-depth interviewing

Few social scientists draw their own probability samples. Secondary data analysts do not draw samples; many others delegate sampling to sampling statisticians, because an error in sampling is financially, temporally, and analytically costly. For those who draw their own probability samples, many universities have inexpensive consulting services to help. To use these services effectively one must understand basic principles of probability sampling and be able to communicate one's research aims.

Small (2009, p. 14), claiming low percentages of probability sampled contactees participate, accuses probability sampling interviewers of burying or omitting their *response rates*. Small claims that because low response rates undermine generalizing from probability samples interviewers should reject probability sampling. Curiously, however, Small implicitly, inexplicably exempts non-probability samples from concern. The chain of reasoning that secures the preference for high response rates, however, makes this exemption untenable.

The motivating links in the reasoning chain are that high response rates mean lower proportions of acquiescent respondents, hence reduced biasing power of volunteers who, owing to social world lumpiness, likely differ in unknown ways from non-volunteers. What analysts seek, therefore, are *not* high response rates; analysts seek to reduce volunteer bias. Response rates are one diagnostic signal on this issue analysts monitor. Thus, Small's (2009, p. 14) denouncement of low probability sample response rates equates to denouncing volunteer-biased samples. Non-probability samples, almost always composed *only* of volunteers, should therefore be rejected. Yet Small accepts non-probability sampling.

Further, contradicting Small's (2009, p. 11) claim that an IDI response rate of 35 % is "highly optimistic," some in-depth interviewers have probability sampled effectively. For example, Mullen (2010) interviewed 50 Yale and 50 Southern Connecticut State University students, obtaining response rates of 81 and 68 %, respectively.

Others claim hidden populations prohibit probability sampling (e.g., Eland-Goossensen et al. 1997). Yet, Rossi et al. (1987) probability sampled sheltered and unsheltered homeless persons in Chicago, reporting response rates of 81 and 94 % respectively; and Kanouse et al. (1999) probability sampled street prostitutes in Los Angeles County, with a 61–89% response rate range.

Surely probability sampling poses challenges, and sometimes it may not be immediately apparent how to proceed. As these exemplars demonstrate, however, early challenge can yield to innovative strategy. Once interviewers accept the challenge, many ostensibly prohibitive situations will, through their creativity and skill, be rendered accessible to probability sampling for IDI.

9 Conclusions

In-depth interviewing is a promising method. Alas, the undeniable ontological condition of social world lumpiness makes respondent selection a serious challenge for any analyst, and in-depth interviewers have almost universally adopted a strategy—non-probability sampling—that ignores the challenge. Thus hobbled by designs that preclude generalizing even as interviewers inevitably generalize, most IDI studies probably betray rather than fulfill the promise of IDI.

Notably, most in-depth interviewers admit one cannot generalize from non-probability samples. Yet, many still maintain such samples provide theoretical insight. Others embrace generalization logics that allegedly justify non-probability sampling. Still others pursue research aims for which non-probability sampling is purportedly sufficient. Underlying each claim is the belief that one's conceptions render research strategies effective.

However, for any conception to validate research operations the conception and its operationalization must align with a plausible ontology, and such alignment cannot occur absent a response to the inescapable partiality of every vantage point owing to social world lumpiness. Social world lumpiness does not smooth to homogeneity simply because a researcher conceives of matters in a particular way; regardless of one's ontological commitments, social world lumpiness invariably destroys the utility of non-probability sampling, whether one's findings are produced through mathematical calculation or otherwise, whether one's generalizations concern substantive or theoretical claims, and whether one's comfort approaching strangers is high or nonexistent. Hence, one must address the threat lumpiness poses.

Current common IDI practice does not address the threat, and most in-depth interviewers have resisted doing so. In his critique of fundamental problems with many statistical analyses, Manski (1995) observed that:

> Empirical researchers usually enjoy learning of positive methodological findings. Particularly pleasing are results showing that conventional assumptions, when combined with available data, imply stronger conclusions than previously recognized. Negative findings are less welcome. Researchers are especially reluctant to learn that, given the available data, some conclusions of interest cannot be drawn unless strong assumptions are invoked. Be that as it may, both positive and negative findings are important to the advancement of science. (p. 3).

We encounter a similar dynamic here, for this paper's message is that an accepted procedure produces almost no knowledge. These negative findings may elicit resistance.

Dismissiveness is unfortunate, but resistance that engages the problems, and perhaps motivates efforts to salvage non-probability sampling, is not unwelcomed. A salvage operation, however, is daunting. In-depth interviewers who sample non-probabilistically cannot simply assert non-probability samples are fine. Instead, they must deploy multiple non-contradictory, ontologically plausible statements to establish that the social world is *not* lumpy, thus implying analysts may study social interaction or meaning-making in one nation (the U.S., say) or community (e.g., suburban mothers) and pronounce with substantive and theoretical specificity about social interaction and meaning-making everywhere. Thus, the social world homogeneity assumption seems untenable.

Non-probability sampling interviewers have two other options. They can establish, with evidence rather than assertion, why the findings above (e.g., the indeterminance of all summary claims when sampling probabilities are unknown, the failure of non-probability samples to reproduce social relations which undercuts excavation of social processes) do not apply to

their study or are generally mistaken. Decades of research validating probability sampling, however, suggest such efforts will fail.

Alternatively, interviewers may avoid *any* and *all* summary claims, limiting their contribution to separately analyzing each respondent as a distinct idiosyncratic instance, only. In other words, if in-depth interviewers hew to the limits non-probability sampling imposes, there is no problem.

Existing studies, however, indicate this third counsel is extremely difficult to honor, perhaps because it contradicts in-depth interviewers' aim as social scientists to summarize respondents' reports and draw broader conclusions. A non-contradictory response, therefore, is for in-depth interviewers to reject non-probability sampling. Rejecting non-probability sampling, and embracing probability sampling, will empower in-depth interviewers to go beyond the existence proof, to reap for all social science the tantalizing bounty of nuanced generalization only in-depth interview studies may possibly provide.

References

Anderson, B.A., Silver, B.D., Abramson, P.R.: The effects of the race of the interviewer on race-related attitudes of black respondents in SRC/CPS National Elections Studies. Public Opin. Q. **52**, 289–324 (1988)

Berg, S: Snowball sampling-I. In: Kotz, S., Read, C.B., Balakrishnan, N., Vidakovic, B. (eds.) Encyclopedia of Statistical Sciences, pp. 7817–7821. Wiley, Hoboken, NJ (2006)

Boas, F.: The limitations of the comparative method of anthropology. Science **4**, 901–908 (1896)

Brunson, R.K: Beyond stop rates: using qualitative methods to examine racially biased policing. In: Rice, S.K., White, M.D. (eds.) Race, Ethnicity, and Policing, pp. 221–238. New York University Press, New York (2010)

Byrne, M.: Sampling for qualitative research. AORN J. **73**, 494–498 (2001)

Campbell, D.T., Stanley, J.C.: Experimental and Quasi-experimental Designs for Research. Houghton Mifflin, Boston (1963)

Das, A., Laumann, E.: How to get valid answers from survey questions: what we learned asking about sexual behavior and the measurement of sexuality. In: Walford, G., Tucker, E., Viswanathan, M. (eds.) The Sage Handbook of Measurement, pp. 9–26. Sage, Thousand Oaks, CA (2010)

Desrosières, A.: The Politics of Large Numbers: A History of Statistical Reasoning, translated by Camille Naish. Harvard University Press, Cambridge, MA (2004)

Eland-Goossensen, M.A., Van De Goor, L.A.M., Vollemans, E.C., Hendriks, V.M., Garretsen, H.F.L.: Snowball sampling applied to opiate addicts outside the treatment system. Addict. Res. **5**, 317–330 (1997)

Feynman, R.P.: Cargo cult science. In: Surely You're Joking, Mr. Feynman: Adventures of a Curious Character, pp. 308–317. Bantam, New York (1985)

Goel, S., Salganik, M.J.: Assessing respondent-driven sampling. Proc. Natl. Acad Sci. USA. **107**, 6743–6747 (2010)

Gomm, R., Hammersley, M., Foster, P.: Case Study Methods: Key Issues, Key Texts. Sage, Thousand Oaks, CA (2000)

Kanouse, D.E., Berry, S.H., Duan, N., Lever, J., Carson, S., Perlman, J.F., Levitan, B.: Drawing a probability sample of female street prostitutes in Los Angeles County. J. Sex Res. **36**, 45–51 (1999)

Kokoska, S., Nevison, C.: Statistical Tables and Formulae. Springer, New York (1989)

Kruskal, W., Mosteller, F.: Representative sampling, IV: the history of the concept in statistics, 1895–1939. Int. Stat. Rev. **48**, 169–195 (1980)

Lincoln, Y.S., Guba, E.G.: The only generalization is: there is no generalization. In: Gomm, R., Hammersley, M., Foster, P. (eds.) Case Study Method: Key Issues, Key Texts, pp. 27–44. Sage, Thousand Oaks, CA [(1979) 2000]

Lucas, S.R.: Theorizing Discrimination in an Era of Contested Prejudice: Discrimination in the United States, vol. 1. Temple University Press, Philadelphia (2008)

Manski, C.F.: Identification Problems in the Social Sciences. Harvard University Press, Cambridge, MA (1995)

Marshall, M.N.: Sampling for qualitative research. Family Pract. **13**, 522–526 (1996)

Merriam, S.B.: What can you tell from an n of 1? Issues of validity and reliability in qualitative research. PAACE J. Lifelong Learn. **4**, 51–60 (1995)

Montemurro, B., McClure, B.: Changing gender norms for alcohol consumption: social drinking and lowered inhibitions at bachelorette parties. Sex Roles **52**, 279–288 (2005)

Mullen, A.L.: Degrees of Inequality: Culture, Class, and Gender in American Higher Education. Johns Hopkins Press, Baltimore (2010)

Neter, J., Wasserman, W., Kutner, M.H.: Applied Linear Regression Models, 2nd edn. McGraw-Hill/Irwin, San Francisco (1989)

Oberfield, Z.W.: Rule following and discretion at government's frontlines: continuity and change during organization socialization. J. Public Admin. Res. Theory **20**, 735–755 (2010)

O'Brien, E.: The Racial Middle: Latinos and Asian Americans Living Beyond the Racial Divide. New York University Press, New York (2008)

Orrange, R.M.: Individualism, family values, and the professional middle class: in-depth interviews with advanced law and MBA students. Sociol. Q. **44**, 451–480 (2003)

Price, S.D.: Infrared astronomy on the Midcourse Space Experiment. Space Sci. Rev. **74**, 81–87 (1995)

Ragin, C.C.: Redesigning Social Inquiry: Fuzzy Sets and Beyond. University of Chicago Press, Chicago (2008)

Rossi, P.H., Wright, J.D., Fisher, G.A., Willis, G.: The urban homeless: estimating composition and size. Science **235**(4794), 1336–1341 (1987)

Salganik, M.J., Heckathorn, D.D.: Sampling and estimation in hidden populations using respondent-driven sampling. Sociol. Methodol. **34**, 193–239 (2004)

Schuman, H., Rieger, C.: Historical analogies, generational effects, and attitudes toward war. Am. Sociol. Rev. **57**, 315–326 (1992)

Small, M.L.: 'How many cases do I need?' On science and the logic of case selection in field-based research. Ethnography **10**, 5–38 (2009)

Suchman, L., Jordan, B.: Interactional troubles in face-to-face survey interviews. J. Am. Stat. Assoc. **85**, 232–241 (1990)

Wald, A.: Sequential method of sampling for deciding between two courses of action. J. Am. Stat. Assoc. **40**, 277–306 (1945)

Wright, K., Copestake, J.: Impact assessment of microfinance using qualitative data: communicating between social scientists and practitioners using the QUIP. J. Int. Dev. **16**, 355–367 (2004)

Wrigley, B.J.: Glass ceiling? *What* glass ceiling? A qualitative study of how women view the glass ceiling in public relations and communications management. J. Public Relat. Res. **14**, 27–55 (2002)

Yin, R.K. : Case Study Research: Design and Methods. Sage, Newbury Park, CA (2009)