

Bakeman, R. & Suma, K. (2021). *Joint Engagement as a Triadic State and Joint Attention as an Infant Skill Shared by Humans and Chimpanzees*. [Peer commentary on the article "[Joint Attention in Human and Chimpanzee Infants in Varied Socio-ecological Contexts](#)" by K. A. Bard, H. Keller, K. M. Ross, B. Hewlett, L. Butler, S. T. Boysen, & T. Matsuzawa]. *Monograph Matters*. Retrieved from <https://monographmatters.srcd.org/2022/03/30/commentary-bakemansuma-86-4/>

## Joint Engagement as a Triadic State and Joint Attention as an Infant Skill Shared by Humans and Chimpanzees

Roger Bakeman  
Georgia State University  
[bakeman@gsu.edu](mailto:bakeman@gsu.edu)

Katharine Suma  
University of Georgia  
[suma@uga.edu](mailto:suma@uga.edu)

The study reported in the monograph by [Kim Bard and colleagues \(2021\)](#) is observational in both senses of the term: it is not experimental and, of greater relevance for these comments, its measurement strategies rely on observational methods. (Full disclosure: RB was the chair of Bard's PhD dissertation committee.) Like all studies, it melds together a goal—addressing a particular research question—with selected samples that are observed in specified settings, guided by an appropriate design, and incorporating conceptually relevant constructs that, in turn, are operationalized with appropriate measurement strategies. The challenge of any study is to assure that all these components are well-chosen and fit together optimally. In that regard, the present study can serve as a model; like a smoothly functioning machine, its thoughtfully selected components work well together, all in the service of addressing a specified research goal.

The central question here is whether joint engagement in one-year-old infants—understood as triadic connectedness involving an infant, a conspecific or conspecifics, and a shared topic—is a common human characteristic and, additionally, whether it also occurs in chimpanzees.

Addressing this broad question, Bard and colleagues argue that the study of joint engagement must be “decolonized.” They define a *joint attention* phenotype as “interactions in which an infant looks or gestures to an adult female to share attention about an object, within a positive emotional atmosphere” (p. 7) and argue that defining joint attention in this way—with its emphasis on a single adult female, objects, and positive emotional tone—reflects research conducted largely with WEIRD samples (i.e., *Western, Educated, Industrialized, Rich, and Democratic*, see Henrich, Heine, & Norenzayan, 2010). The generalizability of any research is limited by its samples, as careful researchers usually note. If conclusions are to apply beyond the home country—to the colonies, as it were—more inclusive and diverse samples are required. And while central conceptual definitions may remain unaltered, operational definitions will likely require modification guided by knowledge of local culture and custom—this is what decolonization entails. And this is what Bard and colleagues deliver.

Their design and samples—observing humans in three different cultures and chimpanzees under three different rearing conditions—fit their question well. Additionally, their expanded definition of *joint engagement*—which keeps the conceptual core of triangular connectedness—is informed by knowledge of variations in behavior, opportunity, and values in the different samples, both human and chimpanzee. Specifically, they cross-classified bouts of joint engagement—defined as an infant engaged with somebody about something—with four sets of mutually exclusive and exhaustive codes:

1. Partner: adult female, adult male, infant, juvenile, multiple, or unclear;
2. Shared Topic: object, food, locomotion, social play, gaze follow, social reference, give and take, or other;
3. Partner’s Emotional Tone: positive, negative, other, unclear, or no emotion; and
4. Initiator: infant initiated joint engagement (IJE), responsive joint engagement (RJE), both RJE and IJE, or unclear/unknown.

They then defined a subset of those bouts—those coded as adult female partner, object topic, positive emotional tone, and infant-initiated joint engagement—as fitting what they characterized as a joint attention phenotype.

On a casual first read, this monograph could be interpreted as an indictment and criticism of joint engagement research with WEIRD samples. We think this may be an over interpretation but that it is nonetheless useful to ask three questions. First, what lessons do we draw for researchers (like ourselves) concerned with joint engagement? Second, what guidance would we offer readers who want to translate the language of this report into the language of other reports concerned with joint engagement? And finally, third, does standard advice concerning observational methods need modification when procedures developed for a particular sample are expanded to diverse samples (e.g., ones that aren’t WEIRD) in which the form (and perhaps function) of central constructs (e.g., joint engagement) may vary? We now address each of these three concerns, but in reverse order.

Bard and colleagues are reaching for more encompassing, less particular, more universal conclusions. Like a Ptolemaic astronomer, our image for this enterprise consists of an almost limitless hierarchical set of nested spheres. Locating ourselves in one sphere and incorporating subordinate spheres, we strive for a larger, more encompassing sphere. The key to our success is finding definitions and procedures that are universal in the more encompassing, superordinate sphere.

We would argue that, in the present context, the conceptual definition of joint engagement—triadic connectedness involving an infant, a conspecific or conspecifics, and a shared topic—constitutes an acceptable universal. There are no, for example, “northern” and “southern” triangles in this sphere, just one triangular arrangement that observers, once appropriately trained and no matter their cultural background, could agree on, no matter the culture of the participants.

We would argue further that the heart of observational methods—defining a set or sets of mutually exclusive and exhaustive codes that characterize a central concept and perhaps its

### 3 Bakeman & Suma

relevant components—likewise constitutes an acceptable universal. Again, there are no “northern” or “southern” versions of this template, just one overarching, universal format that can accommodate a multitude of particulars. As evidence that such conceptual definitions and methodological templates can be used to successfully occupy a more encompassing sphere, we would simply point to the present monograph, which does so with considerable success.

As is often noted, the devil resides in the details; accordingly, we now turn to the terminology and definitions used in the present monograph and suggest how these terms translate to the program of research in which we have collaborated—a program that has been led by Lauren Adamson over the past few decades.

The terms used by Bard and colleagues are joint attention (JA) and joint engagement (JE), with joint attention appearing in their title. In case there is any doubt that these terms are defined and used in a multitude of ways, the monograph authors provide a detailed and helpful table (Table 1) cataloging diverse ways researchers have defined and used these terms. The terms and definitions appear to be shaped by individual researchers’ particular goals, samples, and settings, but—except perhaps for the core triadic definition Bard and colleagues offer—no clear consensus emerges from this table.

Bard and colleagues define joint engagement as superordinate—a bout of joint engagement occurs when an infant is engaged with somebody (not just a mother) about something (not just an object)—whereas joint attention is defined more restrictively. Thus, according to Bard and colleagues, the joint attention phenotype occurs when an infant is engaged with a mother about an object, with the additional restrictions that the emotional tone be positive and that the bout be infant initiated.

The definitions we would give differ, although we admit that, over time, there have been evolutions in the definitions that we and our colleagues use. In addition, we acknowledge that we have not always been completely consistent within our research group. Like Bard and colleagues, we regard joint engagement as more general. Specifically, we think of joint engagement as a state jointly created by infant and other in which they share something—often an object—with the other. This reflects Werner and Kaplan’s (1963) primordial sharing situation (Adamson, 1995; Werner & Kaplan, 1963). In contrast, we think of joint attention more as an infant skill, for which the distinction between infant initiated and infant responded makes sense.

Thus, Bard and colleagues’ definition of joint engagement is similar but somewhat more expansive than ours—and in ways that highlight the importance in any study of matching codes to samples and settings. Bard and colleagues, seeking generality, videorecorded in unstructured situations, whereas in our studies, wanting to provide roughly equal opportunity in different conditions and for different groups, we supplied some structure (e.g., providing toys and limiting partners in Bakeman & Adamson, 1984; and provided various props and suggested activities in Adamson et al., 2009). If we had used Bard’s definitions—thereby allowing for additional partners and topics—we suspect that the effect on our results would likely have been minimal, although it is possible that some of what we called “person engagement,” might

be called “joint engagement” by Bard and colleagues. For example, with food as the topic, nursing could be coded a JE activity if there were an acknowledgment of togetherness, perhaps expressed in mutual gaze or by patting the others person’s skin (the latter an example given by Bard et al.).

In contrast, Bard and colleagues’ definition of joint attention is considerably more restrictive than ours. They do not make the distinction between state and skill that we usually do; what they call joint attention is simply a subset of joint engagement bouts. However, as is the case with our work as well, their definition of joint attention is more infant-centered than is their definition of joint engagement. Their definition of joint attention is even more infant-centered than is ours, requiring as it does that the bout of joint engagement be initiated by the infant.

Bard and colleagues emphasize repeatedly the importance, for their purposes, of observing in everyday contexts, in the infants’ own physical and social ecologies, without constraints imposed by researchers—apart from their introduction of recording equipment into the natural ecology. Their commitment is to observational methods, not tests. Nonetheless, the distinction they apply—between bouts of joint engagement that are infant initiated (IJE) and those to which the infant responded (RJE)—emerges from the testing literature (Mundy et al., 2003). In fact, based on Mundy’s work, we have applied this distinction to tests of our own (Adamson et al., 2021), but have not incorporated it in our codes for engagement state.

Instead, we have made a different distinction, specifically between *supported* and *coordinated* joint engagement. In our research from 1984 on, our definitions of both coordinated and supported joint engagement states preserved the core of joint engagement’s triangular connectedness, but, as we have defined it, coordinated joint engagement requires that the infant explicitly acknowledge the other’s participation, whereas supported joint engagement does not. (Note, in our early work [Bakeman & Adamson, 1984] we used the term “passive joint engagement” rather than “*supported* joint engagement.” The change in modifier was made because we came to believe that the use of the term passive was potentially misleading.)

Of their various samples, Bard and colleagues’ UK sample, which was recorded in the 2000s, is the most comparable to the sample we recorded in the US during the 1980s (both samples were WEIRD). Our setting was somewhat more structured than was theirs. However, if we had applied their definition of joint engagement to our observations (thus combining supported and coordinated joint engagement into a single variable: total joint engagement), our results would change little if at all. Not so joint attention. Perhaps our coordinated joint engagement comes closest to their joint attention but is nonetheless not as restrictive. Coordinated joint engagement requires only infant acknowledgment, not infant initiation, and does not require a partner’s warm emotional tone. Nonetheless, one analytic result is preordained: statistics associated with our total joint engagement will be greater than our coordinated joint engagement, and statistics associated with Bard and colleagues’ joint engagement will be greater than those associated with their joint attention phenotype, by definition.

(Note, direct comparisons of percentages across publications are problematic. Percentages reported by Bakeman and Adamson, 1984, and Adamson et al., 2009, are of time; by Bard and

## 5 Bakeman & Suma

colleagues are of visible 10-s intervals; and by Adamson et al., 2012, and subsequent publications are of mean rating items.)

We now ask, after reading this monograph, what lessons do we draw for own research endeavors, and what advice might we pass on to similar researchers? First might be the importance of explicitly recognizing limitations. Researchers' samples and settings limit generalizability and we all should say so. There is nothing inherently wrong in conducting research solely with WEIRD samples. Often research questions involve comparisons between groups defined by different diagnostic categories, ages, or demographic characteristics within a specified culture—with findings understood to be limited to other similar cultural groups.

Such research becomes problematic when findings are assumed to apply universally to quite different cultural groups or, worse, when findings in one cultural group are considered normative, and differences with other groups are assumed to reveal deficits. Still, even if it is possible to find such worst-case examples in the literature, we should not judge whole literatures by a few bad examples. Increasingly, we suspect, that as awareness evolves—and this monograph can only increase awareness—such examples, if they occur at all, will likely not survive editorial and peer review.

However, when research strives for expanded generality beyond the home country, this monograph serves as a model, demonstrating how samples and definitions need to expand accordingly. It is worth noting—this is our second lesson—that as samples and definitions are decolonized, methods are not. As Bard and colleagues strive toward a more universal, decolonized definition of joint engagement, although their specific codes change, their observational methods do not. This simply demonstrates that empirical methods are not, for example, “northern” or “southern,” but universal, available for use by anyone no matter their culture of origin. For observational methods, it is not the overarching structure of the method but the codes that require understanding of cultural variation, as this monograph demonstrates.

Especially when research strives for generality beyond the home county—and this is our third and final lesson—careful attention to conceptual and operational definitions takes on added importance. If nothing else, reading this monograph has convinced us of the conceptual clarity gained by preserving the distinction between joint engagement as a triadic state and joint attention as an infant skill.

The definition of joint engagement developed in this monograph strikes us as generally useful for the field. It fits literature we know, even though that literature is largely limited to WEIRD samples, and, at the same time, accommodates samples from other cultures as well as samples of chimpanzees experiencing different rearing conditions. Considering partners other than adult females would often have little effect on much of the research we know, in part because samples and settings often limit possible partners. Considering additional topics and settings, however—as this monograph does—could contribute to a useful broadening of studies of joint engagement, although again we recognize that samples and setting often emphasize objects.

The definition of the joint attention phenotype, in contrast, strikes us as less useful because it is restricted in ways that don't reflect the literature we know. Nonetheless, coding partners' emotional tone, or whether bouts of joint engagement were infant initiated, could serve as useful coding adjuncts, just not as part of a core definition.

One final comment: In a sense, the title of this monograph—*Joint Attention in Human and Chimpanzee Infants in Varied Socio-Ecological Contexts*—undersells it. It suggests an exclusive focus on joint attention as an infant skill, whereas we see one of its primary merits as the way it can broaden readers' thinking about and understanding of joint engagement as a state of triadic connectedness involving most generally an infant, a conspecific or conspecifics, and a topic that may or may not involve objects.

### Acknowledgment

It is with a profound sense of gratitude that we acknowledge the intellectual contribution of Lauren Adamson reflected in our commentary. Her death, December 31, 2021, represents a great loss for us and the field

### References

- Adamson, L. B. (1995). *Communication development during infancy*. WCB Brown & Benchmark.
- Adamson, L. B., Bakeman, R., Deckner, D. F., & Nelson, P. B. (2012). Rating parent-child interactions: joint engagement, communication dynamics, and shared topics in autism, Down syndrome, and typical development. *Journal of Autism and Developmental Disorders*, 42, 2622–2635. <https://doi.org/10.1007/s10803-012-1520-1>
- Adamson, L. B., Suma, K., Bakeman, R., Kellerman, A., & Robins, D. L. (2021). Auditory joint attention skills: Development and diagnostic differences during infancy. *Infant Behavior and Development*. <https://doi.org/10.1016/j.infbeh.2021.101560>
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant interaction. *Child Development*, 55, 1278–1289.
- Bard, K. A., Keller, H., Ross K. M., Hewlett, B., Butler, L., Boysen, S. T., & Matsuzawa, T. (2021). Joint attention in human and chimpanzee infants in varied socio-ecological contexts. *Monographs of the Society for Research in Child Development*, 86(4). <https://doi.org/10.1111/mono.12435>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, 33, 61-135. <https://doi.org/10.1017/S0140525X0999152X>
- Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *Early Social Communication Scales (ESCS)*. Coral Gables, FL: University of Miami.
- Werner, H., & Kaplan, B. (1963). *Symbol formation*. Wiley.