

Centers of Attention: a Scaffolded Ecological Information Framework

by Robert P. Levy

Introduction

Research on joint attention in primates and human infants has uncovered a framework of cooperative communication unique to humans: at around 9 months of age, we begin to use gestural signals to share, follow into, and direct what adults attend to (Carpenter, Nagell, Tomasello 1998, Tomasello 2008). These skills suggest an innate disposition to assist others by guiding attentional focus as an effective means toward shared objectives. These behaviors, especially when augmented by tools and repeatable practices for motivating and guiding attention, are referred to as scaffolding. This framework of cooperative communication has received wide attention as an explanation of the origin and present source of the unique human niche, characterized by the ubiquitous exchange of shared intentional practices and narratively curated content.

In light of recent developments in ecological-enactive cognitive science, joint attention-based cooperative communication fits naturally into an account of ecological information, also known as Gibson information, a theory of information as specification by fit between organism and environment. Ecological information can be categorized in terms of four distinct modes of meaningful world-involving encounter, constituting an "ecological information protocol stack".

1. law-based: affordances support direct perception (Gibson 1979, Turvey & Shaw 1992).
2. convention-based: conventionality supports selecting among possible affordances to attend to (Golonka 2015, Golonka & Wilson 2019).
3. scaffolded: scaffolding of joint attention establishes convention-based structured composition of intentional practices (Tomasello 2008) by convention-based mechanisms of semiotic self-control coordinating joint agency (Stjernfelt 2011 & 2014).
4. curated: public symbol systems in the socially shared environment enable content-involving sense-making by socially establishing standards for evaluating scaffolded practices (Wittgenstein 1953, Hutto & Myin 2013 & 2017).

Situated within the broader effort to leverage the ecological approach as a unified theory of intelligent agency, the Skilled Intentionality Framework (SIF) (Rietveld & Kiverstein 2014) has been developed. SIF and related work applying it seeks to extend the definition of affordance to include things like "cultural affordances" so-called (Ramstead, Veissiere, Kirmayer 2016). What we can take from the SIF work is an appreciation of the importance of "solicitations" or task-relevant engagement with a selected ensemble of affordances. However, a modification is needed, as their extended definitions of affordance muddle or conflate the afore-

mentioned distinct modes of information: in order to explain direct perception, "affordance" implies strictly the law-based mode of ecological information that underpins the convention-based mode (Golonka 2015). Ramstead et al's claim of regimes of shared attention and shared intentionality however is appropriate and we develop a narrower definition of "centers of attention" recognizing three basic motive factors of availability to engage, support, and realize the objective of the iterative feedback cycle of cooperative attention to a center, that convention-based information can provide robust scaffolding. We propose and will show that these three motive factors (which are targeted by the distinctive joint attentional skills of sharing, following into, and directing respectively) constitute a powerful and unique kind of joint attention for the reason that they enable agents to coordinate and externally shape Peircean triadic symbols (teleological mechanisms) for self-control on the part of the participants. This mode of scaffolded joint attending, which enables flexible and open ended development of joint agency, differs markedly from animal social coordination on signals that are limited to use by individual agent teleological mechanisms, rather than specifying whole agential modes.

Scaffolding of joint attention is by its very nature an artificial practice, and modern technology offers new prosthetic options with which to conduct it. The emerging "metaverse" of software mediating immersive augmentation of direct and shared experiences offers an opportunity to prosthetically extend the core joint attentional framework of cooperative communication. The proposed architectural pattern of "narrative practice scaffolding" being implemented in the tentative Senters framework, seeks to increase the leverage of scaffolded ecological information by augmenting gestural signaling with respect to instrumented affordance availability threshold events in relation to the motive factors of engaging, supporting, and realizing centers of attention. As such, the familiar means of scaffolding centers of attention can be made legible to computing tools. Software mediated expression of scaffolded ecological information can allow software agents to navigate and support human intention, including social maintenance of propositional content, empowering users to casually and spontaneously curate meaningfully augmented experiences. To those ends, this article initiates an effort to designed a framework of this kind, beginning with a specification that will evolve as implementation proceeds apace.

Law-based Ecological Information

In the second half of the 20th century, James and Eleanor Gibson began the project of developing a phenomenology of perception as prospective biological control guided by affordances of the physical environment. Rejecting the Cartesian view of perception being the result of enriching raw sensory input, they arrived at a theory of perceptual development consistent with the biological understanding of development as differentiation, and of perceptual control systems rather than sensations. This work produced a comprehensive theory of information in terms of the direct fit between agents and law-based dispositions

of the environment. Invariant physical patterns, such as regularities in the array of light reflecting upon surfaces over perspectival movement at the relevant scale, can specify affordances. Affordances, such as a walkable ground, or a climbable tree, specify reliable options to be realized by agents that are able to use them. Effectivities, which are the dispositional properties of agents that complement affordances, can be innately evolved to fit affordances or require differentiation in the course of development to realize the affordances they fit. In ecological psychology the term "task" refers to instances of prospective control-oriented behavior, which is understood to characterize the teleological nature of agents' behavior in general. When an agent actively engaged in a task uses affordances to guide itself toward a preferable state, it is picking up on ecological information. Embodiment of ecological information is the calibration of body and environment in a kind of measured adjustment guided by affordance. To acquire the information supporting perception-action in this way is to measure affordances by the effectivities that fit them.

Convention-based Ecological Information

Affordances have the unique property of supporting direct perception by specifying reliable action opportunities for compatible and prepared agents, but direct perception is not the only way for agents to acquire ecological information. Agents also come to embody ecological information by indirect, convention-based means as well (Golonka 2015). Convention-based information, unlike law-based information, does not guide action, but rather selects, or tips the balance between attending among available affordances of the environment. An example that serves well to elucidate why is that the exterior of a can of beer does not specify the information for getting to the outcome of drinking beer in the same way that the reflectivity of a paved surface specifies the information for slipping and sliding on ice, or the way that the can affords opening by lifting its lever. Instead what the labeling of the can offers is a way of selecting the beer drinking experience, which is then engaged via direct affordances of the can and the liquid. This account of the use of conventions to selectively attend to the direct affordances that must be realized in order to achieve the indicated outcome illustrates that some actionable/perceivable event or feature (eg. a stop sign as object in the layout of the environment) is always needed as a means of selecting some other (eg. controlling one's vehicle in relation to the expected behavior of other cars.)

As Golonka's examples illustrate, the convention-based/law-based distinction is a cross-cutting one, orthogonal to distinctions such as natural versus human-made or social versus nonsocial. The most widely popularized examples of law-based information tend to be natural, non-social examples (such as the affordances of climbable trees to animals that can climb them), but this is mainly for pedagogical reasons due to their relative simplicity. There are human-made, non-social affordances such as steps, ramps, doorways, walkways, and shelters, and natural social affordances such as infant crying and facial expressions, and pointing gestures, which we will explore in more detail. Most social information

is however convention-based, both among humans and animal species. Golonka gives the example of the honeybee waggle dance, which might seem to be a social affordance supporting an ecological law relationship between honeybees and food sources, but is recognized instead as convention-based natural social information because the conditions it is sensitive to may blink in and out of existence. Human gaze-following in contrast is a natural social affordance, because gaze is used in cooperative communication, and it always follows linearly to the immediate location of its target. For comparison, the human deictic convention of composing phrases such as "the room we are in" is conventional, because it can't be used to continuously guide action as a mathematically expressible ecological law the way that gaze can. And finally, it should be clarified that despite the wealth of socially-oriented examples, convention-based information need not be social, as demonstrated by examples such as a domestic cat relying (often unsuccessfully) on the sound of a can-opener to pursue an opportunity for food, or a person deciding between two courses of action based on the outcome of a coin flip.

Scaffolding Ecological Information for Cooperative Communication

Humans are uniquely distinguished in the natural world by our propensity to socially coordinate on shared intentional practices. What is the source of these unique abilities? The most promising answer to this question to date comes from the study of gesture and joint attention. A concrete way in which humans are observably different from other primates is that around 9 months of age, children show a special interest in what adults are attending to, learn to follow what is being attended to, and use pointing gestures to direct the focus of attention. Other primates in contrast demonstrate only rudimentary skills of attention-movement (getting another primate to do something) and attention-getting (calling attention to something the other will respond to in some way) and do not engage shared attention to socially negotiable centers of bearing. What is interesting about human joint attention from an evolutionary perspective is that it seems to constitute a protocol of cooperative communication that can explain the aspects of the human ecological niche that differ so dramatically from other species. Those aspects amount to what we call culture, a kind of narratively constructed augmented reality that we collectively participate in. What seems to have given human populations a comparative advantage is the ability to cooperate better in groups, for example in hunting big game animals.

The ecological approach illuminates a key aspect of the findings of joint attention research: how social information of what to attend to yields the social information of ostensibly why to attend to it. According to the ecological approach, the basic stance of agents is one of prospective control situated in a layout of objects known only by events, encountered by means of prospective control. For the agent, prospective control "requires that one perceive whether the act as a whole is possible, what subacts are possible with respect to the surface layout, and the possible consequences of current subacts if current (kinetic, kinematic) conditions

persist" (Turvey 1992). For humans, the shared information for where to focus attention is most informative in the context of Turvey's three requirements, because it yields the convention-based ecological information of motives for attending. By signaling what to attend to in the immediate context of engaging (option to enter a center of attention), supporting (option to engage a center of attention in support of the present center), or realizing (option to engage a center of attention as a consequence of realizing the present center), centers of attention become social accretions of "when", "how", and "why" information by indicating that attending fulfills an "engaging", "supporting", or "realizing" role respectively, for a given center of attention. I will call this specific kind of convention-based information that joint attention produces "scaffolded ecological information" because it aptly suggests the cooperative context is what makes it useful.

Scaffolding as Coordinated Semiotic Self-control

The congruence of the natural ecological ontology of direct task perception/performance with Tomasello's findings on joint attention perhaps suffices on its own to explain the unique efficacy of human joint attention, however a complementary and further clarifying account comes from Charles S. Peirce's semiotics as elucidated by Frederik Stjernfelt (Stjernfelt 2012, 2014). In the field of biosemiotics, Peirce's triadic theory of symbol has been demonstrating a great deal of utility in making sense of how convention-based functions in forming teleological control (cybernetic) mechanisms in the constitution of biological systems. The core idea from Peirce relevant here is that symbol is a mode of control whereby a sign vehicle is used to arrive at a target, but by means of an "interpretant" or supporting set of dyadic/indexical (or additionally triadic/symbolic) mechanisms, such that the symbol as a whole realizes a generalization or compression in relating interpretant mechanism to targeted end. From a biosemiotic perspective we can view the Gibsonian perception-action feedback cycle as nested within largely innate constitutive system-internal endosemiotic mechanisms of end-directed semiotic control that regulate homeostatic and allostatic reference conditions to produce adaptively favorable outcomes for organisms. Further, biosemiotics helps make sense of animal communication signal systems, whereby there is often complex convention-based agent coordination.

Biosemiotics is relevant to understanding the mechanism of human joint attention because

Curating Scaffolded Ecological Information as Content

We have taken a tour of law-based information for guiding action, convention-based information for selecting action, and scaffolded information for cooperatively selecting action. We have seen that joint attentional scaffolding practices constitute a uniquely expressive mode of interaction supporting sharing of inten-

tional practices. But despite this expressive power, it is still non-representational and content-free. Basic agents do not traffic in facts of how matters in the world stand; fundamentally agents have capacities, develop capacities, and in the case of human scaffolded practices— share capacities. Cooperative communication, even when conducted using language, is not a game of truth but an exchange of skillful capacities, expressed as effective directing of attention (Hutto & Myin 2013). Agents as autonomous control systems embody a basic end-directed unintentionality requisite to the teleosemiotic character of ecological "information for", but do not manipulate contents or "information about".

But given that human actors in the natural world *do* engage in the social exchange and development of information as content, how do we account for that? An answer, perhaps unsurprisingly, begins with scaffolded information. Scaffolding of joint attention does not itself yield content, but given symbol systems in the shared environment, through which the utility of scaffolding may be evaluated in game-like manner, communities of users of such systems curate content, evaluating it on its status of satisfying conditions of success or failure at its function (Hutto & Myin 2017, Wittgenstein 1957). What characterizes content is meta-level attention to the carrying out of scaffolded practices with an eye to the consequences of being reorganized by them (Noë 2015). It is a practice of attending to a public system in which symbols (such as words or pictures) standing for scaffolded episodes, are invoked for comparative consideration of utility. This meta-level practice, itself scaffolded cooperative communication as much as any other, constitutes the curation of content because it allows scaffolded practices to be compared and examined as alternative devices or ways of seeing that are less or more appropriate for objectives of involvement in the world. Claim-making systems can amount to binary truth evaluation, evaluation of relative merit, or other ways of evaluating scaffolded practices construing intentional behavior.

This understanding of scaffolding and curating practices leads to a view of human culture as shared artifice: all cultural artifice scaffolds information for selecting intentional behaviors, but not all cultural artifice is content-bearing. However the achievement of content has so radically transformed what it means to have a self or be a person in a society, that it cannot be overstated how much of what we today call human happens as practices of curating content within public sense-making or claim-making systems. The development of language itself, likely mostly or entirely a human (not innate) achievement, is a perfect example of a tool we constantly use to scaffold attention for purposes of curating content. The same is true for other tools such as diagramming, scorekeeping, musical notation, and computer programming language code. The capacity and reach of such practices stands to be broadly extended by software augmentation of the human protocol of joint attention, as we explore below.

Skilled Intentionality, Empowerment, and Regimes of Shared Attention

We have outlined most of a framework explaining human culture in terms of ecological information: 1. affordances support direct perception, 2. convention-based information supports selection among options of what to attend to, 3. scaffolding of joint attention establishes convention-based structured composition of intentional practices, and 4. public symbol systems in the socially shared environment enable content-involving sense-making by means of evaluating scaffolded practices. However, earlier in discussing law-based and convention-based information we didn't tell the whole story. In addition to understanding prospective control with respect to an given affordance, we also need to know how agents solicit an ensemble of affordances relevant to the task at hand. This question brings to the fore what is arguably the most defining feature of attention, that of prioritization—agents are always under pressure to stay aware of what matters in their surroundings, and must allocate focus wisely as events occur in real time.

The Skilled Intentionality Framework (SIF) (Rietveld & Kiverstein 2014, Bruineberg & Rietveld 2014) addresses precisely this question of how agents engage with multiple affordances together relevantly. The term "solicitation" is used to refer to such a set of affordances that is selected as salient given a setting of intentional control toward some end. Taking a page from Merleau-Ponty, agents are observed to seek out an "optimal grip" on a field of solicitations. Borrowing from the vocabulary of Wittgenstein, the sort of selective openness to a relevant solicitation of affordances is said to be appropriate to a particular "form of life", defined as a relatively stable pattern or norm of intentional behavior. This in a nutshell is the high level program of skilled intentionality.

The account of optimal grip on a field of solicitations offers an enhanced perspective on mechanisms of cooperative communication. Given that agents are seeking to position themselves for maximal leverage, this baseline attentional tendency can assist prosocial behaviors, because it allows agents that are engaged in following the overt motivations of their fellow agents to offer relevant (and not already known to the other) directions that depend on having a broader shared standard of utility. Assuming the Tomasellian framework for joint attentional scaffolding given above, this utility applies equally to pointing out what is available, supporting the presently active target of attention, or realizing the present center of active attention. The result of this is that agents scaffolded by shared centers of attention can explore a topology of relationships among scaffolded centers by maximizing empowerment in addressing each of these three attentional concerns. This mutual exchange, not only of useful scaffolding, but of open-ended flows engaging topologies of possible scaffolded centers of attention, is the hallmark of cooperative communication.

As a final note on the relevance of SIF to our scaffolded ecological information framework (SEIF), there has been a proposal to acknowledge something termed

a "cultural affordance" (Ramstead, Veissiere, Kirmayer 2016) as an explanation of how shared intentional practices are acquired in culture. The idea of "cultural affordance" taken at face value is problematic in its attempt to expand the definition of affordance to something fundamentally incapable of supporting direct guidance of perception/action, however, the intuitive motivation that culturally available information is in some way similar to affordances (despite technically being constituted as convention-based information) is one we find valuable. Affordances following Gibson and later clarified by Turvey and others, are dispositions of the environment specifying law-based information to compatible agents. Loosening this definition (and further involving a dubious classification lumping together notions of convention-based information generically, with what we've termed scaffolded information) makes it harder to talk about real affordances that have the defining property of supporting direct perception. However the cultural affordances paper, leaning on SIF (which we can substitute with EM) proposes a notion of "regimes of shared attention and shared intentionality" which are patterns of practice characterized by feedback loops between joint attention and coordinated intentional practices. At high level of description their formulation of regimes is quite similar to our centers of scaffolded attention. It is our contention that our account of centers as nexuses of iterative collaborative feedback over convention-based selection and composition fulfills the requirements that Ramstead defines for acquisition of the shared forms of life constituting culture.

Cooperative Communication by Coordination of Personal Scaffolding Systems

We wish to specify a framework for involving software in the human ecological niche by means of augmenting the informational modes we've explored here. A natural starting point for a computing framework that extends the innate human framework of cooperative communication is the idea of a personal system for augmenting intentional agency. In the spirit of spirit of "bicycles for the mind" (Licklider 1960, Engelbart 1962, Levy 2018) we will define a "personal attentional scaffolding system" (PASS). As a starting point, a PASS is something like a personal journal where every important event is privately recorded. Many of these events are emitted by tools that we do things with and instruments that register measurements we care about. In order for such a journal to constitute a PASS it must support its owner's activities of soliciting relevant arrays of events in contexts of intentional practice. So we must incorporate some form of agent-like control into the organization of a personal event log for it to be a PASS. The means by which we propose to do so is by deploying attendants (artificial agents) in our personal systems that learn to share, follow into, and direct our attention in keeping with our intentions. Finally, in order to provide attendants with the information required to work within scopes of intention, we need a way to express such scopes, and that is precisely the scaffolded ecological information framework developed above. What is needed therefore is a schematic standard by which to socially share and negotiate attention. So a PASS requires a protocol

for coordination of centers of attention between participants. In summary, a PASS requires:

1. A means of instrumenting and tracking personally and socially relevant events.
2. Attendants that leverage event data and centers of attention that organize it, to augment personal attentional reorganization.
3. A protocol defining an event-sourced schematic standard for collaboratively scaffolding centers of attention.

Of these three components, the protocol requires the most additional definition here. The protocol provides a standard way of sharing gestures of engagement, support, and realization that event-source the scaffolding of centers. These gestures associate instruments that signal threshold conditions (and may have effects) with the gestured phase of attending to a center. This provides attendant processes with the structured task-relevant information needed to present relevant guidance to human users and narratively organize software experiences. For the detailed technical specification of the senters protocol and PASS framework, see <https://github.com/senters/senters/blob/master/senters-technical-spec.org>.

Narrative Practice Scaffolding as a Paradigm of Situated Computing

We propose Senters as specific implementation of what we see as a more generic architectural pattern for augmenting the human joint attentional framework of cooperative communication, which we call "narrative practice scaffolding" (NPS). As the name implies, it targets the full ecological information protocol stack, from narratively curating content down to instrumenting directly engaged affordances.

Perhaps the most important differentiating feature of NPS as an approach to computing is that it proposes users should be able to tell their own stories in using software. And these stories should be as or more functionally important to the operation of how software mediates real-world activities than what the developers of systems and instruments being leveraged had in mind. What is more, it should often be possible to conduct narrative reorganization of system behavior casually and spontaneously in situ. Shared stories that shape shared attention and control of systems is the main goal we have our sights on in this work. We want to add a functionally effective narrative layer to existing technology stacks, for functionally tractable content curation.

This article is a work in progress, like the Senters project itself. The contents of this article will be subject to refinement as systems of the kind described here are actually developed and tested. At the time of this writing no system has been created fulfilling the plans laid here, but the exercise of laying out the design will guide the work that is to come. That work, including its situated use in real-world scenarios, will serve both a practical or engineering goal of creating very flexible and open-ended socially situated composition of computing

experiences, and a theoretical or scientific goal of exploring the scaffolded mind hypothesis and scaffolded ecological information.

References

- "Philosophical Investigations" Wittgenstein 1953 <https://static1.squarespace.com/static/54889e73e4b0a2c1f9891289/t/564b61a4e4b04eca59c4d232/1447780772744/Ludwig.Wittgenstein.-.Philosophical.Investigations.pdf>
- "Man-Computer Symbiosis" Licklider 1960 <http://worrydream.com/refs/Licklider%20-%20Man-Computer%20Symbiosis.pdf>
- "Augmenting Human Intellect: A Conceptual Framework" Engelbart 1962 <http://dougengelbart.org/content/view/138>
- "The Senses Considered as Perceptual Systems" Gibson 1966 <https://b-ok.cc/book/3554998/9b8962>
- "The Ecological Approach to Visual Perception" Gibson 1979 <http://b-ok.cc/book/864226/e0dd92>
- "Affordances and Prospective Control: An Outline of the Ontology" Turvey 1992 <https://canvas.brown.edu/courses/755816/files/20990546/download?wrap=1>
- "Social cognition, joint attention, and communicative competence from 9 to 15 months of age." Carpenter, Nagell, Tomasello 1998 <http://booksc.xyz/book/15451904/9e3c39>
- "An Ecological Approach to Perceptual Learning and Development" Gibson & Pick 2000 <https://b-ok.cc/book/879525/7d1f69>
- "Origins of Human Communication" Tomasello 2008 <http://b-ok.cc/book/541274/39859f>
- "Empowerment: A Universal Agent-Centric Measure of Control" Klyubin, Polani, Nehaniv 2005 https://www.researchgate.net/publication/4201381_Empowerment_A_Universal_Agent-Centric_Measure_of_Control
- "A free energy principle for the brain" Friston, Kilner, Harrison 2006 <https://www.fil.ion.ucl.ac.uk/~karl/A%20free%20energy%20principle%20for%20the%20brain.pdf>
- "On Intelligence From First Principles: Guidelines for Inquiry Into the Hypothesis of Physical Intelligence (PI)" Turvey & Carello 2012 <https://booksc.xyz/book/42939095/6bac90>
- "The Evolution of Semiotic Self-Control" Frederik Stjernfelt 2012 <https://www.researchgate.net/profile/Frederik-Stjernfelt-2>
- "Radicalizing Enactivism" Hutto & Myin 2013 <http://b-ok.cc/book/2554656/1b7ea8>
- "Natural Propositions" Frederik Stjernfelt 2014 https://books.google.com/books/about/Natural_Propositions.html?id=X37ioQEACAAJ
- "A Rich Landscape of Affordances" Rietveld & Kiverstein 2014 <https://www.tandfonline.com/doi/full/10.1080/10407413.2014.958035?scroll=top&needAccess=true>
- "Self-organization, free energy minimization, and optimal grip on a field of affordances" Bruineberg & Rietveld 2014 <https://www.frontiersin.org/>

- articles/10.3389/fnhum.2014.00599/full
- "Laws and Conventions in Language-Related Behaviors" Golonka 2015 <http://booksc.xyz/book/50082310/1e8631>
 - "Strange Tools: Art and Human Nature" Noë 2015 <http://b-ok.cc/book/2640649/b1b44d>
 - "Cultural Affordances: Scaffolding Local Worlds Through Shared Intentionality and Regimes of Attention" Ramstead, Veissière, Kirmayer 2016 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4960915/>
 - "Evolving Enactivism" Hutto & Myin 2017 <http://b-ok.cc/book/2947353/09d772>
 - "Attentional actions – an ecological-enactive account of utterances of concrete words" Jasper C. Van Den Herik 2018 <https://sciendo.com/article/10.2478/plc-2018-0005>
 - "Free energy, empowerment, and predictive information compared" Biehl, Guckelsberger, Salge, Smith, Polani 2018 https://www.mis.mpg.de/fileadmin/pdf/abstract_gso18_3300.pdf
 - "Getting into predictive processing's great guessing game: Bootstrap heaven or hell?" Hutto 2018 <https://philpapers.org/rec/HUTGIP>
 - "Steps to an Ecology of Bicycles for the Mind: A Situated Programming Manifesto" Levy 2018 <http://sinters.info/situated-programming>
 - "Joint Attention for Multi-Agent Coordination and Social Learning" Dennis Lee, Natasha Jaques, Chase Kew, Douglas Eck, Dale Schuurmans, Aleksandra Faust 2021 <https://arxiv.org/abs/2104.07750?context=cs.MA>