

Epistemic Interaction

Alan Dix

Computational Foundry Swansea University

1 November 2020

<https://alandix.com/academic/papers/epistemic-2020/>

There is a growing expectation that forms of machine learning or other data-driven algorithms will enter virtually every aspect of human–computer interactions. In many examples the user’s interactions in the physical or digital world are monitored in order to modify subsequent system behaviour. Usually this is done with minimal or no change to the primary interactions, you act as normal, but these actions are passively sensed and analysed. However, there is the potential for minor shifts in interaction that can substantially increase the information available for system adaptation. That is, we can deliberately design interactions to increase the knowledge available to the system – I call this *epistemic interaction*.

In the rest of this article we’ll put this concept in context of human–object and human–human communication. We are used to the idea of designing devices and systems so that users have better understanding of what they are doing through, visualisations, feedback and affordances. However, in human–human collaboration we constantly shape our actions so that they offer subtle often implicit cues to others. In *epistemic interaction*, we extend this to human–system interactions.

Before working through this more theoretical account, we’ll consider a concrete example of epistemic interaction.

First example – An intelligent heating control

A traditional thermostatic heating control panel has a temperature knob that is twisted to select the desired temperature. Digital controllers often substitute the dial for +/– or up down arrow buttons to increase or decrease the target temperature.

A well-known problem with this sort of control is that many people confuse the rate of warming and target temperature. For example, if the target temperature is 20 degrees and the room is only 15, the user, feeling cold, might increase the target temperature further even though the heating is already increasing the temperature. This is a reasonable misunderstanding as non-thermostatic systems often have similar looking controls where the dial is about the amount of heat being produced.

This confusion has the potential of becoming even more pronounced in an intelligent system that may proactively change the target room temperature based on, say, the user's activity or time of day.

Imagine instead a controller that just as before has +/- or up/down button, however these are not about setting the target temperature but instead mean "I'm too cold/hot". Pressing the '+' button means "I want it warmer than it is now" and maybe double tapping it means "a lot warmer". If the intelligent system is already increasing the temperature, then the user's input is effectively ignored, although would reinforce the system's learnt rules. If the system had thought the room was warm enough, then the '+' would tell it both to make it hotter now and to learn this for later.

Note how the small difference in the interaction technique changes the information available to the system to make adaptations. This is *epistemic interaction* in practice.

Human-object interactions – by nature

Affordances in the natural world

Within HCI, *affordances* are the most well-known aspect Gibson's ecological understanding of perception [Gi79], the way in which the perceptible properties of objects in a sense announce their potential for human action. Of course, in a human-object interaction the objects are passive and it is our perceptual systems that have evolved to immediately grasp this action potential; for example that this stone is small enough to hold and yet large enough break open a nut.

Of course the world we live in now is full of human-made artefacts. A hammer, car or website rollover has not been around for the hundreds of thousands of years of human development. Sometimes these do recruit natural affordances, for example the size of a hammer shaft suggests it is hand graspable. However this does not in itself establish its full action potential, in particular its intimate connection to the nail, or ultimately joining wood together.

Affordance seeking in the constructed world

Happily, we are naturally *affordance-seeking* creatures, constantly learning the patterns around us even of radically new classes of thing, so that we can establish new forms of perceptions-action relationships [GP03].

Early electric light switches recruited natural affordances: they are finger sized and project out from the back plate so suggesting they can be manipulated by a fingertip. However, the relationship to turning on and off a remote device or light is learnt. Having learnt this, later electric light switches can be more subtle and the idea of a switch or button to be pressed to make an effect is borrowed into computer buttons and toggles, and these themselves have become part of young children's cultural background, so that the idea coloured patch

that you touch is now part of many young children's implicit understanding before they are ever tall enough to reach a light switch.

Epistemic action

Another key term in ecological psychology is *epistemic action*. In both familiar and unfamiliar situations, if the perception of an object or space is in some way ambiguous or incomplete, we perform some sort of physical action in order to gain more knowledge. This may be a shift our body position, perhaps moving our eyes or head; it may be a movement in space, such as looking round the corner to seek out a landmark; or it may include manipulating an object, for example picking up a hammer to assess its weight.

Note both affordance and epistemic action speak about the intimate relationship between perception and action. Affordance says that perception is about seeking out action and epistemic action says that action may be used to enhance perception.

Human–system interactions – passive sensing

Sometimes intelligence or AI is very upfront in a design, for example in a medical decision support system. However, in many of the examples the adaptations and interventions are more subtle and crucially the sensing is often passive. For example, while you shop on a web site your choices are monitored and then later this may be used to offer suggestions. Similarly activity recognition systems may use vision techniques or IOT-enabled objects to enable context-sensitive services.

I coined the term '*incidental interaction*' for situations where the user's actions on a *primary task* are sensed and the knowledge and learning from that may be used by the system to modify its behaviour on some subsequent *secondary task* [Dx02, Dx17].

Sometimes the primary and secondary task are very closely related, for example an intelligent cookery aide watching you add ingredients and prompting you if you appear to have forgotten a stage from the recipe [BA14, KW18]. In other examples the secondary task may be more indirectly related, for example in the Tiree internet-connected shop-open sign, the act of turning on the illuminated 'open' sign is used by the system to change the display of web information [Dx17].

Crucially, these systems usually passively sense the user's actions. The system may try to learn new patterns, but the primary task is designed to be as optimal as possible for its own purposes. It is the job of the system to make meaning from the interactions.

In some ways the intelligent system is behaving like the affordance-seeking human encountering new kinds of objects.

Human–object interactions – by design

Norman's adoption of affordance took a design perspective on human–object interactions when the objects are themselves human-made artefacts [No13]. Rather than rely on the user to have to learn an adapt to potentially arbitrary representations, Norman suggested that designers should deliberately seek to understand the existing perceptual affordances (termed *signifiers* in Norman's later work [No08]) of the user whether innate or learnt, and then deliberately design systems so that their perceptual characteristics recruit the users existing palette of affordances.

When this was first introduced in the late 1980s, the primary source of existing affordances were physical phones, such as electric light switch and physical buttons. However, now many will themselves be form our digital culture. For example if dsigning interactions for a large-scale public display or AR system one might choose mid-air swiping actions that mimic those for in-contact swiping on a mobile phone and make this affordance perceptually visible through design elements on the distant large display or overlaid AR that look similar to those on the phone interface.

Looking back it is clear that artefacts in the constructed world are not arbitrary relying totally n human's affordance and meaning seeking nature, but are already in forms that recruit natural or existing cultural affordances. In some cases, this may be because those that did not fell into disuse; in other cases, it is due to slow gradual co-evolution of artefact design and cultural affordances over many generations; and, in some cases, this may be the skill of inspired crafts-folk.

Norman and Gaver's early work on affordances [No13 ,Ga91] both brought this understanding to the user interface design world and made it *explicit*. That is, we deliberately design visual and interactive elements of systems so that they offer appropriate perceptual information to the user.

Human–human interactions – cooperation through action

Much of the earliest work on human communication focused on the explicit channels in the spoken or written word (e.g. grammar, semantics) and other symbol systems (e.g. semiotics). Later work highlighted focused on the many side-channels in everyday conversation from tone of voice to facial expression and eye movement and the way these amplify or modify the raw words. Similarly, Austin's [Au62] and Searle's [Se69] speech act theory of uncovered the many layers of meaning that lie within and yet beyond the plain content of speech and also, the crucial *performative* role of *communication as action*, doing things in and of itself.

Notions of *repair* in conversational analysis [FD94] and Clark and Brennan's *common ground* [CB91] go beyond the single statements of speech and shows how the patterns of conversation continually act to both carry forward a primary topic, but at the same time monitor and offer feedback and conformations of understanding. For example, when asked "when does it leave?" when standing at the Swansea bus stop, one might answer "the

Swansea bus leaves at ten past the hour”, implicitly offering confirmation of the contextual understanding of the question (the Swansea bus, not some other bus or a train) and thus the potential, but only if needed, for the interlocutor to correct the interpretation if incorrect.

Analysis of collaborative situations have established that this shaping to allow interpretation goes well beyond direct communication.

When physically moving a large object, such as furniture, one might explicitly say things like “can you lift your end a bit?”, but much of the communication is felt through one’s mutual actions on the object – *feedthrough* [DX94], if you push to the left, I feel it and move slightly to my right. Furthermore, you might explicitly exaggerate your movements in order that they are more easily interpreted – onomatopoeic action [Dx09, DG21].

Ethnographies of collaboration have shown that this is also true when the collaboration is not so direct. Notably Heath and Luff’s classic study of the London Underground control room [HL92] demonstrated the importance of subtle cues such as half-overheard telephone conversations or the way someone was looking towards the large display enabled controllers to modify their behaviour or prepare for future more direct interactions even if they were not consciously aware of it.

One of the key concepts of ethnomethodology is the notion of *accountability* [Ga67], the way in which actions are “put together as publicly observable, reportable occur” [BS98]. That is in human–human interactions we are constantly shaping the way we perform an action in order to make it comprehensible to others.

Epistemic interaction – sense-able by design

Epistemic interaction is a natural extension of this human–human communication principle. Even if the most optimal way to carry out a task makes it invisible to others, we may choose, implicitly or explicitly, to adapt our performance so that it is more apparent. Can we do the same for human–system collaboration adapting interactions so that they offer more information to the system to make inferences about our goals, beliefs and behaviours?

Taking another parallel, the use of affordances in HCI encourages us to *design* system appearances and behaviours so that their, otherwise invisible, action potential is apparent to users. Can we explicitly *design* interactions that expose otherwise hidden information and this allow systems to learn better?

You can probably think of examples where this is done with *explicit* additional user interactions. For example after reading help text one is often asked to give a thumbs up/down to say whether the information presented was useful. If this is optional then many users simply skip the feedback hence reducing or adding bias to the relevance feedback. If this is made mandatory or hard to ignore, then it has the danger of adding friction to the interface, reducing user engagement and damaging user experience.

Looking back to the notions of *primary* and *secondary* task from incidental interaction. The relevance feedback is likely to be of use to aid future interactions, possibly for others (secondary task), but the user is being asked to put in additional work now (primary task).

The most successful incidental interactions are where the primary task is not explicitly interrupted, but is sensed in order to improve future interaction. For example, where past purchasing behaviour (implicit positive relevance) is used to produce recommendations. Advertisers use click-through rate in a similar way.

The design challenge of epistemic interaction is to make small changes to the primary task, that do not add noticeably to the perceived effort, but which act to increase the information available for adaptations.

A-B testing can be seen in this light. Small changes are made to the system and the user's behaviour is logged. The variant that has better outcomes is then chosen as the preferred long-term design choice. By definition, there is an expectation that one of the variants will produce slightly worse user outcomes (in terms of efficiency or experience), but the variants are usually close enough that users will be utterly unaware (e.g. pixel-level placement), or doesn't care.

Arguably A–B testing is self-referential as the information revealed is precisely to make the choice, but it does emphasise that epistemic interaction is not just possible, but widely practiced.

In some ways this has connections with the expected, sensed, and desired framework [BS05] for (primarily physical) sensing-based systems. In the ESD framework one explores what is currently *sensed* about the object by the system, what is the *expected* behaviour when an action is made on the physical object and any other interactions that are *desired*. The framework creates the space to think about gaps in the sensing (prompted often by expected behaviour) and gaps in actional potential (desired) that could possibly be delivered by sensed (or potentially) sensed actions that are not currently mapped to a system response. The domain of ESD is like that of implicit interaction [Sc00] where the sensing is primarily about enabling or contextually modifying actions in the primary task.

For an epistemic action we are typically (but not solely) concerned with information for a secondary task; however this could be the same as the primary task at a future date, as in the intelligent hearing example.

Operationalising epistemic action has two aspects:

- a ***design challenge*** – matching alternative possible interactions to information that would be useful for future interaction adaptations or applications
- a ***selection criterion*** – using this fit and the additional value of the information to weigh against other criteria

For the first one needs to consider different potential interactions and what can be sensed under each alternative and also think about the potential purposes or needs for information.

One can then assess the extent to which the sensed data can feed into different kinds of information, subject of course to the usual assessment for privacy and consent.

For the second the value of the information gained by each alternative needs to be set alongside other criteria such as user experience, interaction efficiency, computational cost and development effort. Clearly if there are strong advantages for one alternative on user-centred criteria this would be chosen, but where alternatives are close or uncertain, the information gathering potential can be used as a decider.

The intention is to have an end interaction that is a viable and reasonable one for the primary task, but which as a side effect (incidentally) delivers useful information.

One potential problem about such interactions is that the lack of an explicit information gathering action means the user will be less aware that actions are being monitored, however, being largely invisible at the moment of gathering does not mean that this should not be upfront in the overall system design and that the information gathered is not available for scrutiny. In some cases this is a minimal concern, for example where the learning and adaptations purely influence the users own home. However, where the information is for broader learning and interaction by others, for example the use of learning analytics to improve future students' studies, more explicit consent may be required.

Example – scrolling vs accordion information displays

As a second practical example, let's consider the choice of interaction technique for showing selected snippets of information in a help system or search results. Various documents have been selected using an algorithm and are available in the form of title, paragraph length snippet (~10 lines) and a link to the full document.

Several design alternatives are being considered:

1. a standard search results listing the title the title of each and a link to the full document, possibly with the snippet at the top of the full-document page
2. a long scrolling page with headings, snippets and 'more' link for each
3. the same but as accordion where each title has an open/close icon
4. variant of (3) where at most one page is open at a time (previous open section is closed as a new one is opened)

An advantage of (1) is that standard web analytics can be used to measure, for example, click-through and time spent on each page. This can help assess the actual relevance of the information and the *information scent* of the titles [CP01]. However, the navigation between pages may become annoying for some users, especially on a mobile device with poor signal compared with a single page design.

We might therefore reject design (1), but considering it has led us to think about its advantages and hence the potential for gathering implicit relevance feedback.

Let's say option 2 comes out best in terms of actual user interaction based on a user study or maybe designer's intuition. It is easiest to rapidly scroll up and down the display spotting things of interest. It is possible to study such scrolling patterns from detailed interaction traces [Dx16], but this is hard to do and fairly uncertain, especially as an automated exercise.

However, in options (3) and (4) we have easy to harvest information, not unlike the click-through data from (1). Opening an accordion section says "this looks interesting" the time spent before closing it or opening another section gives a measure of the interest of the underlying snippet and then clickthrough and dwell time on the full document can tell us how good the snippet paragraph was as information scent.

Option (4) is slightly more informative than (3) as with option (3) as a user or (3) might first of all rapidly open several potentially interesting titles, and then scroll up and down them. The trace of opening might then look as though only the last opened title had actually ended up being of value.

Table 1 summarises the various criteria and the order of options. While option (2) offers the best immediate user experience, other options, particularly (3) and (4) offer better information. Of course, this information can be used to improve future behaviour and hence future user experience.

Table 1 – criteria for selecting result display format

Criterion	Measured by	Option order	Comment
User experience	Expert analysis or user study	2 > 3 > 4 > 1	
Information scent of title	Click through to relevant snippets/docs	1 > 4 > 3 > 2	For (2) impossible to distinguish scent of title and snippet
Sufficiency of snippet	Long dwell on snippet, but no click-through	4 > 3 > 2 > 1	Could also mean confusing but irrelevant
Information scent of snippet paragraph	Click through to relevant document after reading the snippet	4 = 3 = 2 > 1	
Relevance of document	Dwell time on final document	All equal	

If the differences in user experience between (2) and (3) are small, then one might prefer (3) to (2) as a final choice. Of course, one can always obtain finer grained understanding of the difference in terms of user experience between the options using A-B testing!

Note too that the user experience differences between the options may also depend on the device and situation. For example, (4) might be a little annoying on a large screen where there is ample screen real estate to open multiple results to read and compare, but on a

small-screen mobile device it may be no worse or even than (3). That is, one might use the differences between devices to make small differences in the interface so that additional information is gained from some that might help the same user, or another user, on a different device at a different time.

References

- [Au62] J. Austin: *How to Do Things With Words*. Cambridge (Mass.) 1962, paperback: Harvard University Press, 2nd edition, 2005, ISBN 0-674-41152-8.
- [BS05] S. Benford, H. Schnädelbach, B. Koleva, R. Anastasi, C. Greenhalgh, T. Rodden, J. Green, A. Ghali, T. Pridmore, B. Gaver, A. Boucher, B. Walker, S. Pennington, A. Schmidt, H. Gellersen, and A. Steed. 2005. Expected, sensed, and desired: A framework for designing sensing-based interaction. *ACM Trans. Computer–Human Interaction*, 12(1):3–30. DOI: 10.1145/1057237.1057239
- [BA14] R. Blasco, Á. Marco, R. Casas, D. Cirujano and R. Picking (2014). A Smart Kitchen for Ambient Assisted Living. *Sensors*, 14(1):1629-1653. DOI:10.3390/s140101629
- [BS98] Graham Button and Wes Sharrock (1998). The Organizational Accountability of Technological Work. *Social Studies of Science*: 28(1):73–102. DOI: 10.1177/030631298028001003
- [CP01] S. Card, P. Pirolli, M. Van Der Wege, J. Morrison, R. Reeder, P. Schraedley, and J. Boshart (2001). Information scent as a driver of Web behavior graphs: results of a protocol analysis method for Web usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. Association for Computing Machinery, New York, NY, USA, 498–505. DOI: 10.1145/365024.365331
- [CB91] Clark, Herbert H.; Brennan, Susan E. (1991), Grounding in Communication. Chapter 7 in *Perspectives on socially shared cognition*, Resnick, L. B.; Levine, J. M. (eds.), American Psychological Association, ISBN 1-55798-376-3
- [Dx09] A. Dix (2009). *Language and Action (2): from observation to communication*. Posted on May 18, 2009. <https://alandix.com/blog/2009/05/18/language-and-action-2-from-observation-to-communication/>
- [DG21] A. Dix , S.Gill, D. Ramduny-Ellis and J. Hare (2021). Section 15.7 The origins of language in *TouchIT*. OUP.
- [DX94] A. J. Dix (1994). Computer-supported cooperative work - a framework. In *Design Issues in CSCW* Eds. D. Rosenburg and C. Hutchison. Springer Verlag. 23-37.
- [Dx02] A. Dix (2002). beyond intention – pushing boundaries with incidental interaction. *Proceedings of Building Bridges: Interdisciplinary Context-Sensitive Computing*, Glasgow University, 9 Sept 2002
- [Dx16] A. Dix (2016). Challenge and Potential of Fine Grain, Cross-Institutional Learning Data. *L@S: Third Annual ACM Conference on Learning at Scale*, ACM, pp. 261–264. DOI: 10.1145/2876034.2893429
- [Dx17] Dix A. (2017) Activity Modelling for Low-Intention Interaction. In: Weyers B., Bowen J., Dix A., Palanque P. (eds) *The Handbook of Formal Methods in Human-Computer*

- Interaction*. Human–Computer Interaction Series. Springer, Cham. DOI: 10.1007/978-3-319-51838-1_7
- [FD94] D. Frohlich, P. Drew and A. Monk (1994). Management of Repair in Human-Computer Interaction. *Human–Computer Interaction*, 9(3–4):385–425
- [Ga67] H. Garfinkel, *Studies in Ethnomethodology* (Englewood Cliffs, NJ: Prentice-Hall, 1967).
- [Ga91] Gaver, W. Technology affordances. In *Proceedings of CHI 91*. ACM Press: NY (1991), 79-84.
- [Gi79] Gibson, J. J. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin (1979).
- [GP03] Gibson, E. J. and Pick, A. D. *An Ecological Approach to Perceptual Learning and Development*. Cary, NC: Oxford University Press (2003).
- [HL92] Heath, C., Luff, P. (1992). Collaboration and control: Crisis management and multimedia technology in London Underground Line Control Rooms . *Computer Supported Cooperative Work* 1:69–94. DOI: 10.1007/BF00752451
- [KW18] T. Kosch, P. Woźniak, E. Brady, and A. Schmidt (2018). Smart Kitchens for People with Cognitive Impairments: A Qualitative Study of Design Requirements. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Paper 271, 1–12. DOI: 10.1145/3173574.3173845
- [No08] Norman, D. Signifiers, not affordances. *Interactions*, 15, 6 (2008).
- [No13] Norman, Donald A. (2013): *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.
- [Sc00] Schmidt, A. Implicit human computer interaction through context. *Personal Technologies* 4, 191–199 (2000). DOI: 10.1007/BF01324126
- [Se69] J. Searle (1969). *Speech Acts*, Cambridge University Press, ISBN 0-521-09626-X.