

Chapter 2 - Human Needs and Search Behaviour

How to learn from queries and behaviour without inventing the person behind them

A search system records three queries from the same browser:

boiler pressure keeps dropping
combi boiler pressure drops no visible leak
boiler repair tenant landlord responsibility

The record also contains two result clicks, several measured intervals and no booking attributed to the browser. That is enough information to suggest a story.

Perhaps a tenant noticed the pressure falling, opened two unhelpful pages and then searched for the person who should pay for the repair. When no clear answer appeared, the tenant gave up. An SEO report could describe this as a failed journey and recommend a page that moves people more quickly towards a booking.

Nothing in that account is absurd. The problem is that the record does not establish it.

The browser may belong to a homeowner who was learning enough to speak intelligently to an engineer. The first page may have been useful and supplied the phrase *no visible leak*. The final results page may have answered the tenancy question directly, so no external click was necessary. A telephone call may have followed after the number was copied. Someone may have been searching for a relative. Several people may have shared the device. The measured interval may include careful reading, confusion, a conversation, a background tab or activity the search provider never observed.

Each explanation fits the recorded events. They do not support the same conclusion about what the person needed, whether the results helped or what the organisation should change.

This is the problem at the centre of Chapter 2. SEO practitioners work with queries, impressions, clicks, visits, calls, transactions and customer records because those observations can reveal valuable patterns. Yet the decision they want to make usually concerns something larger: a problem, a person, a **task**, an unmet need or an outcome. The practitioner must reason from the retained trace towards that wider explanation without silently adding facts that were never observed.

The alternative is not to declare human behaviour unknowable. It is to separate what happened from what was recorded, identify the explanations that remain possible and seek evidence capable of distinguishing the explanations that matter. That discipline turns incomplete data into bounded knowledge rather than an overconfident story.

One browser record can support several human stories

Begin with what the boiler record really contains. Three requests were associated with the same browser identifier under the search system's processing rules. Two qualifying selections were recorded. Some elapsed times can be calculated between recorded events. The available attribution process found no booking connected to that identifier.

Even this plain description depends on definitions. We need to know what counted as a **query**, which results counted as impressions, what event produced a click, how the browser was recognised, how long the provider retained the identifier and which bookings were available to the attribution system. The record can be accurate under all those rules while still representing only part of the activity.

The first missing element is identity. A browser identifier is not automatically one person. It can be reset, shared or used across several tasks. One person can also use several browsers, devices, applications and offline channels. The three queries may belong to one person, several people or one person acting on behalf of somebody else.

The second missing element is circumstance. The record does not say whether the heating has failed, whether the property is rented, whether the searcher has money for a repair, whether the problem feels dangerous or whether trusted help is available. Those facts could alter both the words submitted and the sensible next action.

The third missing element is meaning. A click establishes a recorded selection, not what the person thought before selecting it or understood afterwards. An interval establishes a duration between events visible to the measuring system, not continuous attention. The missing booking says that no matching booking was attributed within the observed route. It does not prove that the person did nothing, that the search failed or that the wider problem remained unresolved.

We can now name the boundary we have crossed. The human activity happened in the world. Instrumentation captured selected events and produced an **interaction record**. When an analyst says that the person was confused, satisfied or ready to buy, the analyst has moved beyond the record and made an **inference**.

Inference is not a defect. Every useful analysis contains it. The defect is allowing an inference to masquerade as an observation. Search-log research has long required analysts to distinguish the interactions a system records from the needs, motives and outcomes it does not directly contain.^{[1](#)}

That distinction improves the question. Instead of asking, "What was the user's intent?", we can ask, "Which human **situations** and tasks could have produced this trace, which of them would change our decision and what evidence could distinguish them?" To answer well, we need to separate the different things that the convenient word *intent* has been hiding.

The human problem is larger than the query

Imagine that the home in the boiler case is becoming colder. The person may be a tenant who is unsure whether touching the appliance is safe. Money may be limited. The landlord may be difficult to contact. A child or older relative may be in the property. None of these details is the query, but each could change what help is useful.

Those circumstances form the person's **situation**: the relevant features of time, place, role, prior knowledge, constraints, available channels and social setting. Situation is not a complete psychological portrait or a marketing segment. It is an analytical selection of the circumstances that might affect the problem. An analytics row does not contain the complete situation, and a researcher must decide which parts to investigate.

Within that situation, the person may be trying to bring about an outcome. That desired change is a **goal**. Reliable heating might be the main goal. Avoiding danger, understanding responsibility and preventing an unnecessary bill may be connected goals. The organisation may value a completed repair booking, but its conversion event is not automatically the person's goal. The person could achieve safe heating through a landlord, an existing engineer or another route that creates no booking for the measured business.

Reaching the goal requires activity. A **task** is an organised piece of work or responsibility undertaken towards an outcome. In this case it may include recognising the fault, deciding whether it is urgent, identifying who should act and arranging help. Those activities can be treated as one task or several subtasks. The boundary may be set by the person, a

landlord, an employer, a researcher or an analytics rule. Tasks can pause, resume, overlap, be delegated or contain conflicting goals.

The person may then require informational help. An **information need** is a provisional explanation of what information could allow progress in the situation or task. The person may need to recognise warning signs, distinguish likely causes, verify a tenancy rule or locate qualified support. Calling this a provisional explanation matters. The need is not a hidden sentence waiting to be extracted from the query. It is an account of the gap between the present situation and useful progress.

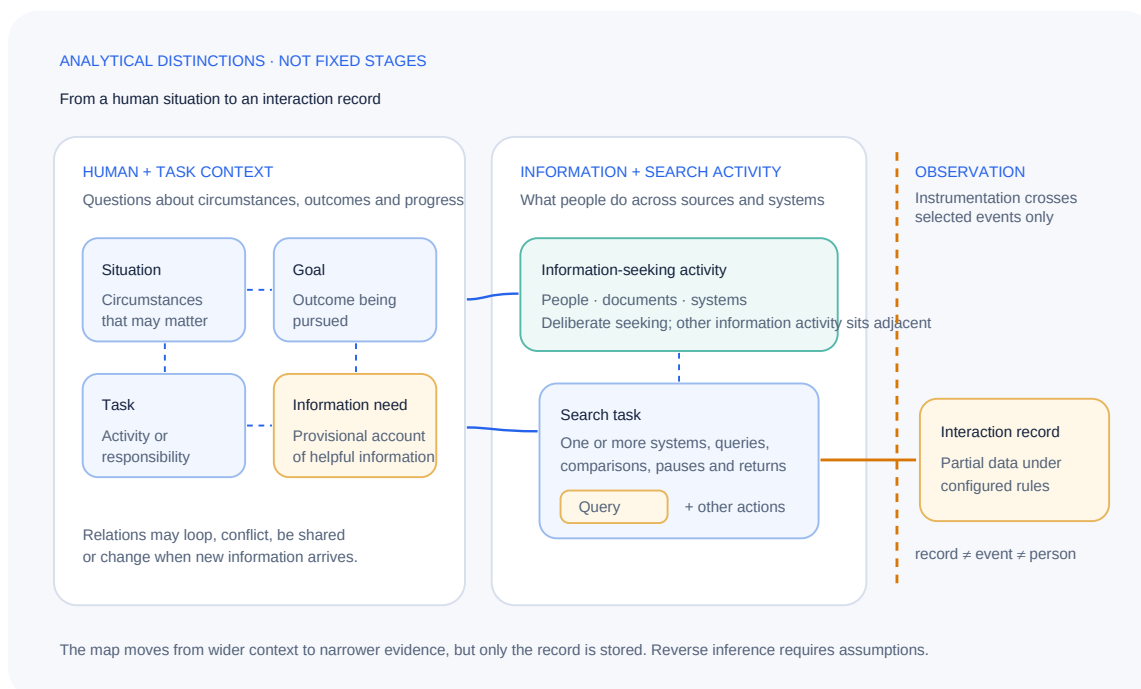
Information can be sought from more than a search engine. Asking a landlord, checking the boiler manual, telephoning an engineer and consulting a neighbour are all forms of **information-seeking activity** when they are deliberate attempts to obtain information. Useful information may also be encountered without deliberate seeking, perhaps when a relative mentions a warning sign. That encounter can alter the task or create a need the person had not recognised.

A **search task** is the bounded part of the wider activity in which one or more search or retrieval systems are used. It can contain several queries, result examinations, comparisons, pauses and returns. It may cross a public search engine, a government site, a landlord portal and a specialist directory. Its boundary is an analytical claim unless the person or task setting supplies and confirms it.

A **query** is narrower still. It is an explicit request submitted to a search system at a particular moment. Typed words, spoken words and structured requests can all qualify. A query is one move within search activity. It is evidence of what was submitted, not a complete account of why it was submitted. If the system rewrites the query, adds location or generates a follow-up request, those machine-produced requests must remain distinct from the person's own words.

Finally, instrumentation turns selected system events into the interaction record with which we began. One action can create several records, one imperfect record or no retained record. Filtering, identifiers, client and server coverage, consent, attribution and processing determine what survives.

These distinctions can be arranged as a map:



The construct map separates human and task context, information and search activity, and the observation layer. Situation, goal, task and provisional need influence one another. Search actions cross an observation boundary before they become interaction records. These are analytical distinctions, not fixed stages.

The map is not an eight-stage journey. A task can be imposed before the person adopts a goal. New information can change the perceived situation, goal, task, vocabulary and standard of success. Searching can be collaborative or

delegated. Several queries can serve one stable task, and one query can serve several subgoals. The relationships run in more than one direction.

Nor does the map make every funnel illegitimate. A funnel can summarise a deliberately chosen sequence of recorded stages, such as impression, visit and completed booking. It can be useful for locating an operational loss. It becomes misleading only when those stages are presented as the complete human process. A funnel of records is still a funnel of records.

The practical improvement is immediate. "The user's intent was to book a boiler repair" becomes "this browser submitted repair-related requests; diagnosis, responsibility, reassurance and arranging help remain plausible purposes." The second sentence is more cautious, but it is also more actionable. It reveals the paths the content may need to support and the evidence still missing.

One question remains. Are we assuming that a complete need exists before the query and merely fails to fit inside it, or can the need itself become clearer while the person searches?

An information need can be clear, uncertain or negotiated

Sometimes the answer is simple. A person may know the exact manual they want and search for its title. They may need a known telephone number or a fixed fact. The immediate target is stable, the vocabulary is adequate and a short query is an efficient request to a capable system.

Other problems are difficult to express because the person is searching partly to understand what the problem is. Robert Taylor described a movement between an unexpressed concern and the formal question eventually presented to an information system. The lasting value of his work is the warning that the submitted request has already been transformed by the person's understanding of the system and the language available to them.² It is not evidence that every searcher passes through four measurable stages.

The word *compromise* is useful here. A system-facing question may be adapted to what the searcher thinks the system will accept. That adaptation is not automatically poor. *Boiler pressure keeps dropping* may be concise because the person expects the engine to infer common causes. A long natural-language question could still omit cost, responsibility or fear. Query length therefore cannot serve as a scale of how clearly the wider problem is understood.

Nicholas Belkin and colleagues offered a structural reason for this difficulty. If the problem concerns something the person does not yet understand, precise specification may be hard because the missing knowledge is exactly what would be needed to formulate the best request. Their anomalous-state-of-knowledge account and early interactive system showed that fuller problem descriptions contained structure missed by conventional best-match queries in that setting.³ The work does not establish that every need is a private cognitive anomaly or that an interview recovers a correct hidden state.

Thomas Wilson widens the frame again. People rarely seek information merely to possess more information. They may be trying to achieve safety, permission, reassurance, practical action or social confirmation.⁴ Whether they search depends on cost, available sources, consequences and barriers as well as on recognising that something is not known. The tenant may need a landlord to act more than another article to read.

This leads to a real disagreement about what an information need is. Some accounts treat uncertainty or a problematic state of knowledge within the person as important, even if it cannot be observed directly. Situated accounts place the problem in a particular moment and ask what gap, help or obstruction the person constructs while trying to move forward. Brenda Dervin's Sense-Making work is influential here.⁵ Dialogical work goes further by showing that questions can be formed through conversation, available documents and institutional expectations. Anna Lundh, for

example, studied pupils negotiating research questions with a librarian rather than simply reporting a finished inner need.⁶

The chapter does not need to choose one theory for every case. It needs the operational consequence. Queries, interviews, workshops and survey answers are all produced under conditions. An interview can reveal motives, meanings and off-platform events that a log cannot see. It can also be shaped by memory, wording, social expectations and the conversation itself. A workshop may help a participant articulate a problem while simultaneously changing the participant's understanding of it.

For SEO, the phrase "we discovered the true intent" is therefore usually too strong. A better claim names the method and scope: participants in these event-cued interviews described these goals; these query sequences are consistent with these task explanations; these support conversations repeatedly exposed this obstacle. Such language is not timid. It tells the next decision-maker exactly what can be reused.

Once a need can be stable, uncertain or formed through interaction, a later query cannot be interpreted merely as a clearer version of a fixed original request. Searching may preserve the target, but it may also change the person's vocabulary, understanding or task.

Searching can preserve a target or change the problem

Consider a person searching for a named landlord portal. They might try several spellings, follow a poor result and return to the same target. The queries change, yet the sought item remains stable. Repetition in this case does not make the task exploratory.

Now return to the boiler sequence. The phrase *no visible leak* may have been present in the person's understanding from the start. It may also have been learned from the first repair guide. The move to *tenant landlord responsibility* may be a new subtask created by reading, an old subtask expressed later, a correction after irrelevant results or a request introduced by another person.

The changed query establishes only one fact: the submitted request changed. It does not carry its own explanation.

Several research traditions help separate the possibilities. Gary Marchionini distinguishes lookup from learning and investigation. Lookup includes finding a known item or checking a fact. Learning involves accumulating and interpreting knowledge. Investigation requires more open comparison, evaluation or synthesis.⁷ These activities can overlap, but the distinction prevents a long exploratory model from being forced onto a routine lookup.

Marcia Bates' berrypicking model describes activity in which useful pieces accumulate across sources while new information changes the query and opens new directions.⁸ Carol Kuhlthau's work shows that extended knowledge-construction tasks can initially increase uncertainty before a clearer focus emerges.⁹ Pertti Vakkari followed students through a research task and connected changes in their knowledge with changing vocabulary, tactics and relevance judgements.¹⁰ Allen Foster's naturalistic work offers a necessary correction to orderly stages: opening, orientation and consolidation can recur, overlap and loop.¹¹

The names matter less than the different questions they answer. Bates explains how useful pieces can accumulate. Kuhlthau makes constructive uncertainty visible. Vakkari connects task stage and developing knowledge with behaviour in a narrow longitudinal setting. Foster prevents those patterns from becoming compulsory order. Marchionini keeps stable lookup available. None tells us how often every form occurs across all contemporary searches.

Manual studies of longer query sequences have identified specification, generalisation, branching, recurrence and multitasking.¹² Large-scale models can also learn that rapid, closely related re-querying predicts labelled

dissatisfaction under particular conditions. Such findings change the probability of an explanation. They do not make the pattern identical to its label.

More queries may indicate struggle, but they may also indicate productive comparison or learning. Fewer queries may indicate an excellent direct answer, but they may also indicate an inaccessible interface, a trusted off-platform alternative, fatigue or abandonment. Stopping, satisfaction, correctness and completion are not the same event.

The appropriate interpretation depends partly on what the person could realistically do. That brings the wider situation back into the analysis.

Context changes which actions are possible

Search behaviour is often explained by attaching a label to a person: expert, urgent, mobile, multilingual, high intent. The label then appears to predict the behaviour. This is appealing because labels fit neatly into rows. It is weak reasoning when the mechanism between the condition and the action remains unstated.

Context matters because it changes the feasible action set. A condition may alter which words are available, which sources can be reached, how much effort an action requires, whether privacy is possible and where the task can be completed. The same person can therefore behave differently across tasks, devices, languages and social settings without acquiring a new fixed persona.

Task complexity illustrates the point. Studies across field, naturalistic and controlled settings connect task properties with differences in query formulation, source use, time and number of moves.¹³ Yet *complexity* may refer to the number of subtasks, uncertainty, cognitive demand, expected product or perceived difficulty. A long sequence can result from a complicated task, diligent comparison, poor results or an inaccessible interface. A complex task can produce a short trace if the person asks a trusted expert, delegates or stops.

Prior knowledge changes the resources available to the searcher. Familiarity can supply specialist vocabulary, known sources and criteria for judging evidence. A large historical toolbar study found different within-domain patterns among people classified as experts, with smaller differences outside their apparent field.¹⁴ But expertise and success were inferred through behavioural proxies. The study supports a scoped association and a plausible mechanism. It does not allow an SEO tool to identify experts from long queries, technical language or visits to specialist sites.

Time pressure is not the same as perceived risk. An experimenter-imposed five-minute limit compressed result inspection and changed query rate in one laboratory study.¹⁵ That shows a mechanism under the tested conditions. It is not a measurement of chronic pressure or a household emergency. Research on perceived risk has found contradictory relationships with information search because prior knowledge, social expectations, available help, avoidance and substitute strategies can all change the response.¹⁶ A short high-stakes trace is not automatically careless or commercially ready.

The physical and social environment can create the need as well as constrain the response. A diary study in the 2007 mobile environment found that location, activity, time and conversation repeatedly shaped information needs.¹⁷ Device and interface then change input, viewport, connection, privacy, interruption and completion routes. Historical feature-phone and desktop comparisons show that interfaces can shape queries.¹⁸ Their numerical differences are not current smartphone baselines.

Language is another resource, not merely a demographic label. Multilingual searchers may change query or result language according to topic, proficiency, perceived relevance, trust and interface presentation.¹⁹ Country or language group does not by itself identify the cause of a difference. Content availability, platform familiarity, recruitment and translation can be entangled. A UK investigation cannot treat one English query as proof that English is the only useful language for that person.

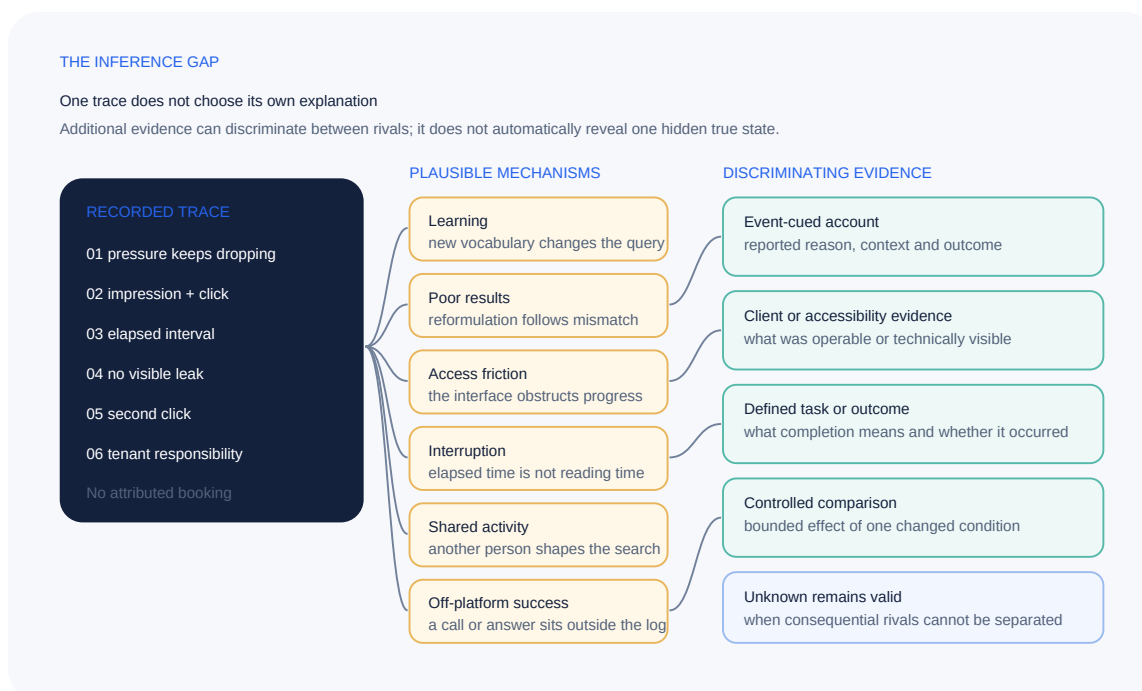
Accessibility barriers can create traces that are easy to misread. Research with blind and low-vision participants found repetition, backtracking, delay and impulsive action among coping tactics used when a system was difficult to operate.²⁰ Those specific patterns are not a diagnostic signature for disability. Functional needs, assistive technology, experience, content and interface interact. Standards evaluation and work with varied disabled users answer complementary questions.²¹ One automated score, participant or screen reader cannot stand for every access condition.

Searching can also be shared. People may generate terms together, divide result review, send findings or search on behalf of someone else.²² An account or browser does not establish one autonomous need owner. Yet proxy search should not become the default explanation either. It is a possibility to preserve until the evidence identifies it.

Finally, absence from the data does not prove absence of need. Access, affordability, permission, time, skill, familiarity, autonomy and human help can keep a problem outside the observed route. Performance studies have shown substantial differences in operational and strategic Internet skills even among people already online.²³ The opposite explanations remain possible: there may be no need, the answer may already be known or another route may simply be preferred.

These conditions interact. Expertise changes how task difficulty is experienced. Language choice changes with topic and interface. Accessibility depends on the combination of functional need, technology and content. Device is entangled with location, activity and network. The useful question is not "which persona is this?" It is "which condition could change the feasible action or the meaning of the trace, and what evidence would distinguish that mechanism?"

The inference problem can now be shown directly:



One recorded search trace can be explained by learning, poor results, accessibility friction, interruption, collaboration or successful off-platform action. Additional evidence can rule explanations in or out, but it does not automatically reveal one hidden true state.

Once these rival mechanisms are visible, familiar behavioural metrics become easier to handle. They are not mysterious signals with fixed psychological meanings. They are operational records whose interpretation depends on definitions, context and comparison.

Behavioural records do not explain themselves

The central distinction is between an event, the operational measure created from it and the human construct an analyst hopes to understand. Skipping a link in that chain is how a number becomes an unsupported conclusion.

An impression is not examination, and a click is not relevance

An **impression** is recorded under a platform's rules for serving, rendering or eligibility. It may not establish that the result entered the visible viewport, received attention, was read or was remembered. **Examination** is a separate construct. Eye tracking can provide evidence about visual fixation in a controlled setting, but a fixation does not prove comprehension and visual evidence does not represent every non-visual interaction.

A **click** records a qualifying selection inside a designed choice environment. The choice depends on more than relevance. Controlled experiments show that position can change click probability even when independent relevance judgements favour another result.²⁴²⁵ Presentation also matters. In particular studies, bold query matches changed selection after rank and judged relevance were controlled, while longer snippets helped assigned informational tasks and hindered navigational tasks.²⁶²⁷

Raw click-through rate is therefore not an absolute relevance measure. The selected result may be relevant, prominent, familiar, attractively phrased or surrounded by weak alternatives. The unclicked result may never have been examined.

Clicks are still useful. Repeated selections can support estimates of relative preference when exposure, examination, comparison and validation target are specified. The safe conclusion remains at the level the design supports. "This result received more qualifying selections under these conditions" does not become "this page satisfied the need" without additional evidence.

It also does not become a ranking-factor claim. Nothing in the interaction studies cited here establishes that a current search engine uses a site's click-through rate, dwell time or another behavioural metric as an organic ranking signal. Evidence about how people choose results is not automatically evidence about the private production system that ranks them.

Dwell time is a definition before it is a duration

The phrase *dwell time* sounds like one clock. In practice, different systems can measure different intervals.

Server dwell often begins with a result click and ends with the next event visible to the search provider. It can include activity that occurred elsewhere and may remain open after the final observed click. **Client dwell** uses browser or page-lifecycle events. It can still count a background tab, distraction or open-but-unread time. **Trail dwell** includes downstream pages and answers a wider question about the path after the initial result.²⁸

The measurement must be defined before it can be interpreted. A fourteen-week naturalistic study found no simple general relationship between display time and judged usefulness, with substantial variation by person and task.²⁹ Later segmented models predicted labelled satisfaction better than a single threshold.³⁰ That improvement shows that conditions matter; it does not give every duration one stable human meaning.

In the boiler case, a short interval could be a fast answer, an accidental click, a technical failure or an obvious mismatch. A long interval could be careful reading, difficult language, confusion, interruption or an inaccessible page. Page length, reading burden, device, task and subsequent action all matter. "Long is good" is not a simplified expert rule. It is an unsupported assumption.

No click is not the same as success or failure

A **no-click query** is an event pattern: under the platform's rules, no qualifying external-result selection was recorded. **Abandonment** requires an additional operational rule deciding that the query or session ended. Neither term explains why.

Research on *good abandonment* established that some no-click searches can be satisfied by information shown on the results page.³¹ Direct prompting has also found satisfied, dissatisfied and neutral reasons among abandoned searches.³² Modern answer-rich interfaces make useful no-click behaviour increasingly plausible. The same record can still arise from poor results, interruption, reformulation, a decision not to act or off-platform continuation.

This exposes a conflict hidden by the word success. The search provider may value interaction satisfaction. The publisher may value a visit. A researcher may care whether the information was correct. The person may care whether the real-world task was completed safely. The organisation may care about a booking or avoided support call. These outcomes can align, but they are not one measure.

A direct answer can satisfy the searcher while reducing publisher traffic. A click can benefit the publisher while sending the person to misleading information. A conversion can occur without full understanding. A correct page can be inaccessible. A fact-finding study that separated query quality, result identification, answer extraction and verification found that assessors still could not determine many outcomes from rich natural trails.³³ **Unknown** is sometimes the most accurate result.

A session is made by an analyst

Analytics systems often group events using an identifier and an inactivity threshold. The resulting **session** can be a useful and reproducible operational unit. It is not a naturally observed boundary around one human goal.

Manually annotated query sequences show interleaved and hierarchical goals. One task may resume after a long pause while another begins seconds after the previous event.³⁴ A fixed timeout can therefore split one task and merge different tasks. Cross-device activity, shared devices, sign-in changes, cookie deletion and movement between search, apps, calls and offline action add more uncertainty.

The unit should fit the decision. A query supports analysis of one submitted expression. A click chain supports analysis of navigation after an exposure. A timeout session supports an aggregate defined by that timeout. An inferred task or customer journey supports a wider question only when its identity and boundary assumptions are stated. Calling a row a person or journey does not make those assumptions true.

We now have a reason to add evidence, but not a reason to collect everything. The missing evidence should be chosen by the human construct and decision at stake.

Evidence must be chosen for the question

No method provides an unrestricted view of the human event. Each opens one window and closes another.

Transaction logs offer scale, sequence and repeated observation inside the instrumented platform. They omit many client events, off-platform actions, motives, knowledge and explanations of success. Client instrumentation can add viewport, scroll, touch and lifecycle events while still failing to prove attention or understanding. Eye tracking observes overt visual fixation in a tested setting, not comprehension or natural population behaviour.

Self-report and interviews can elicit felt urgency, confidence, reasons and remembered outcomes that no click contains. Those accounts can be affected by recall, wording, nonresponse and social expectations. A preregistered meta-analysis found substantial discrepancies between logged and self-reported digital-media quantities.³⁵ The result does not make the log the truth about motive. It shows that reported duration and instrumented duration are different measures.

Diaries and experience sampling can capture needs close to the situation, including needs that never become searches. Observation and think-aloud can expose barriers and tactics while changing the activity being observed. Laboratory experiments can identify a bounded causal contrast under assigned conditions, while artificial tasks and recruited populations may limit transfer. Operational systems can accurately record a call, booking or completed job without revealing why it occurred, whether search caused it or whether the person's wider goal was served.

Large datasets do not remove these distinctions. Digital-trace research separates **representation error**, concerning who and what entered the platform window, from **measurement error**, concerning how activity became records and constructs.³⁶ A recent three-country web-tracking study found substantial device and browser undercoverage in its

panels and used simulations to show how incomplete capture could bias estimates.³⁷ Its numerical results should not be transferred to another analytics stack. Its mechanism is the durable lesson: a precise count from an incomplete window can still misrepresent person-level behaviour.

Begin with the decision, then build the evidence chain

Suppose the team asks whether the boiler guide should be rewritten. Opening the engagement dashboard first would allow available metrics to define the problem. A stronger investigation begins by stating the decision and the outcome that matters.

The relevant construct might be safe progress towards resolving pressure loss, comprehension of tenancy responsibility or qualified access to help. The team must then identify the population and situations to which its conclusion should apply. It asks who and what could enter the dataset, which event or answer represents the construct, what the instrumentation captured, what processing changed and which rival explanations remain compatible. Finally, it asks whether evidence from the studied setting transfers to the intended task, language, device and market.

The complete reasoning order is:

Decision > Construct > Population > Coverage > Operationalisation > Instrumentation > Processing > Rival explanations > Transfer

This chain is an academy synthesis grounded in interactive information-retrieval evaluation, log methodology and digital-trace error frameworks.¹³⁶³⁸ It is not a demand that every low-risk wording change receive an elaborate research programme. The depth of investigation should follow the consequence of error and the explanations that must be distinguished.

Triangulation is designed, not accumulated

Adding a second method is useful when it contributes evidence about a different construct or error process. A search log can identify an event to discuss. An event-cued interview can add the participant's reported reason, surrounding circumstances and off-platform action.³⁹ Accessibility evaluation can reveal a barrier that click data cannot diagnose. A lawful call record can establish a downstream event. A controlled comparison can test whether a bounded page change altered an outcome.

The methods may still share error. A log and interview can involve the same selected population and platform window. Their units may be incompatible: an account-level trace, a person-level account and an organisation-level conversion total cannot be merged casually. Agreement may reflect shared assumptions. Disagreement may occur because the methods observe different objects and should be examined rather than averaged away.

The target construct determines the useful evidence. If the question concerns examination, client or attention evidence is relevant. If it concerns whether the guide solved the problem, a click and duration need a defined outcome or person-level account. If it concerns the causal effect of presentation, a credible comparison is required. If it concerns why people failed, qualitative inquiry supplies information absent from ordinary logs. Simulated work tasks can support a controlled question only when they are credible and relevant to the intended participants.⁴⁰

The absence of an unmediated view is not a reason to abandon measurement. It is a reason to make the inference inspectable. Once the important rivals and evidence limits are visible, the team can choose an action proportionate to what is actually known.

From an ambiguous trace to a responsible SEO decision

Return to the question: should the boiler guide be rewritten?

The trace cannot prove that one tenant wanted to book a repair and failed. It also does not prevent the team from improving the guide. Several rival explanations may support the same reversible change for different reasons. A person learning diagnostic vocabulary, a tenant looking for responsibility and a homeowner preparing to call an engineer could all benefit from clearer pathways, provided the underlying information is accurate and the design remains usable.

The reasoning can be recorded in seven connected steps.

1. State the decision. Decide whether this guide should change, for which audience and towards which outcome.

"Analyse engagement" is not a decision. "Should the guide make diagnosis, tenancy responsibility and qualified repair routes easier to distinguish?" is.

2. Name the constructs. The team may care about recognising urgent warning signs, understanding responsibility and reaching suitable help. It may also care about qualified repair enquiries. These human and owner outcomes can support one another without becoming identical.

3. Describe the actual records. The known evidence is a browser-linked query and click sequence plus the absence of an attributed booking. Provider, event definitions, identity unit, time period, platform window and processing must remain attached. The evidence is not yet a verified person journey.

4. Preserve rival explanations. The changes in query may reflect learning, poor results, better wording, task decomposition, interleaving or a changed actor. The final no-click may reflect a direct answer, interruption, dissatisfaction or off-platform action. Language, accessibility and available channels may change the route.

5. Request evidence that discriminates. Event-cued interviews could provide situated accounts of the changes.

Accessibility work could identify interaction barriers. Client instrumentation could clarify whether a page remained displayed, though not whether it was understood. Lawfully connected call or booking records could establish selected outcomes. A controlled copy change could test a bounded presentation decision.

6. Match action to evidence and risk. If several explanations support clearer routing, the team can make that reversible improvement without declaring one motive proven. Any instruction about boiler safety requires the appropriate technical expertise and authorised review. Behavioural telemetry does not qualify an SEO practitioner to create a high-stakes factual answer.

7. Record unknowns and reversal conditions. The decision should state what remains unresolved, what observation would change the recommendation and which population, interface or instrumentation change requires revalidation. Unknown is a valid result when consequential alternatives cannot yet be distinguished.

The protocol does not produce a confidence score. It produces an inspectable decision. Another practitioner can see which events were recorded, which meanings were inferred, why the action was considered proportionate and what would cause the team to change course.

The content response also becomes more useful. If people are acquiring vocabulary, the page may need progressive explanation rather than repeated exact-match phrases. If one query serves several wider tasks, the page may need clear routes for diagnosis, responsibility and repair rather than one forced intent. If a direct answer resolves the search, publisher traffic and user benefit may diverge. If backtracking reflects an accessibility barrier, changing the keyword target will not solve the problem.

The same structure matters for Valinor's future decision layer. A record such as `high_dwell` cannot safely sit beside a fixed recommendation such as `keep_page`. The system must retain the event definition, observation window, candidate constructs, rival mechanisms, applicability, evidence class, risk, unknowns and the bounded action actually justified. Chapter 2 does not decide the graph, retrieval system, model hierarchy or scoring architecture. It establishes the distinctions that any later architecture must preserve.

The discipline this chapter adds

A query is neither useless nor the need itself. It is a recorded request made at one point within a larger human activity. Situation, goal, task, information need, information-seeking activity, search task and query answer different questions. The interaction record belongs beyond an observation boundary and never becomes the person merely because it contains precise fields.

Some searches preserve a stable target. Others change as people learn, acquire language, compare sources, divide the task or involve someone else. Context changes the actions that are possible without creating universal behavioural signatures. Impressions, clicks, dwell, no-click patterns and sessions can supply valuable evidence when their operational definitions, coverage and rivals remain attached.

Expert analysis therefore rejects two easy positions. It does not treat behavioural data as a transparent account of motive, and it does not dismiss the data because it is incomplete. It states what was recorded, identifies the construct needed for the decision, preserves plausible alternatives and requests evidence capable of distinguishing the alternatives that matter.

Chapter 3 will use this discipline to examine query meaning, inferred intent, reformulation and journeys in greater detail. Later chapters will handle instrumentation, identity, customer tracking, attribution and causal experiments. They begin from the same boundary: a human event, a system action and a retained record are connected, but they are not the same thing.

Source notes

The notes below identify the evidence used in the chapter and, where necessary, the limits on what each source can establish. The numbering matches the links in the text.

Human situations, needs and information seeking

1. **Jansen, B. J. (2006).** "Search Log Analysis: What It Is, What's Been Done, How to Do It." Methodological review of what logs record, omit and require analysts to define. [Author PDF](#)
2. **Taylor, R. S. (1968).** "Question-negotiation and information seeking in libraries." Exploratory basis for distinguishing a problem from its system-facing expression; not a universal stage test. [Paper](#)
3. **Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982).** "ASK for information retrieval: Part I." Early interactive-retrieval theory and design study about problems that resist precise specification. [DOI](#)
4. **Wilson, T. D. (1981).** "On user studies and information needs." Theoretical account placing information seeking within wider human needs, roles and environments. [Author archive](#)
5. **Dervin, B. (1998).** "Sense-making theory and practice." Overview of a situated approach to gaps, helps and movement through circumstances. [DOI](#)
6. **Lundh, A. (2010).** "Studying information needs as question-negotiations in an educational context." A dialogical challenge to treating need as a recoverable private object; illustrated in a narrow school setting. [Open article](#)
7. **Marchionini, G. (2006).** "Exploratory Search: From Finding to Understanding." Conceptual distinction among overlapping lookup, learning and investigation activities. [DOI](#)
8. **Bates, M. J. (1989).** "The Design of Browsing and Berrypicking Techniques for the Online Search Interface." Influential model of evolving queries and accumulating useful pieces across sources. [Author version](#)
9. **Kuhlthau, C. C. (1991).** "Inside the Search Process." Mixed-study synthesis of thoughts, actions and feelings in extended knowledge-construction tasks. [DOI](#)
10. **Vakkari, P. (2001).** "A Theory of the Task-Based Information Retrieval Process." Longitudinal evidence on changing vocabulary, tactics and relevance judgements in eleven students' research tasks. [DOI](#)
11. **Foster, A. (2004).** "A Nonlinear Model of Information-Seeking Behavior." Interview-based naturalistic study with 45 interdisciplinary academics. [DOI](#)

12. Rieh, S. Y., & Xie, H. (2006). "Analysis of Multiple Query Reformulations on the Web." Manual analysis of varied reformulation patterns in purposively long historical sessions. [DOI](#)

Context, language and access

13. Byström, K., & Järvelin, K. (1995); Li, Y., & Belkin, N. J. (2008). Field evidence on task complexity and a faceted framework for task source, goal, action and product. [Byström & Järvelin](#) - [Li & Belkin](#)
14. White, R. W., Dumais, S. T., & Teevan, J. (2009). "Characterizing the Influence of Domain Expertise on Web Search Behavior." Large historical toolbar-log study using inferred expertise and success proxies. [Microsoft Research](#)
15. Crescenzi, A., Kelly, D., & Azzopardi, L. (2015). "Time Pressure and System Delays in Information Search." Randomised laboratory time-limit study with 43 participants. [DOI](#)
16. Gemünden, H. G. (1985). "Perceived Risk and Information Search." Meta-analysis reporting extensive contradiction in the proposed risk-search relationship. [DOI](#)
17. Church, K., & Smyth, B. (2009). "Understanding the Intent Behind Mobile Information Needs." Four-week diary study with 20 participants in the 2007 mobile environment. [DOI](#)
18. Kamvar, M., & Baluja, S. (2006). "A Large Scale Study of Wireless Search Behavior." Historical comparison of feature-phone, PDA and desktop requests; mechanism evidence, not a smartphone baseline. [Google Research](#)
19. Steichen, B., et al. (2021). "How Do Multilingual Users Search?" Five studies of query and result-language choice. [DOI](#)
20. Vigo, M., & Harper, S. (2013). "Coping Tactics Employed by Visually Disabled Users on the Web." In-situ evidence from 24 blind and low-vision users. [DOI](#)
21. W3C Web Accessibility Initiative. "Involving Users in Evaluating Web Accessibility." Current authoritative guidance on combining conformance evaluation with varied user involvement. [W3C guidance](#)
22. Morris, M. R. (2008). "A Survey of Collaborative Web Search Practices." Self-reported practices among 204 knowledge workers at one technology company. [Microsoft Research](#)
23. Hargittai, E. (2002); van Deursen, A., & van Dijk, J. (2011). Performance evidence on heterogeneity in online operational, information and strategic skills in specific historical populations. [Hargittai](#) - [van Deursen & van Dijk](#)

Interaction records and interpretation

24. Joachims, T., et al. (2005). "Accurately Interpreting Clickthrough Data as Implicit Feedback." Controlled eye-tracking and rank-manipulation study showing examination and trust effects. [Cornell PDF](#)
25. Craswell, N., et al. (2008). "An Experimental Comparison of Click Position-Bias Models." Large live ranking intervention showing position effects and rank-specific model limits. [Microsoft Research](#)
26. Yue, Y., Patel, R., & Roehrig, H. (2010). "Beyond Position Bias." Live adjacent-result experiment separating selected presentation effects from rank and judged relevance. [Google PDF](#)
27. Cutrell, E., & Guan, Z. (2007). "What Are You Looking For?" Eye-tracking experiment in which snippet length helped informational tasks and hindered navigational tasks. [Author PDF](#)
28. Kim, Y., et al. (2014). "Comparing Client and Server Dwell Time Estimates." Demonstrates that server, client and trail clocks are different measurements. [Microsoft PDF](#)
29. Kelly, D., & Belkin, N. J. (2004). "Display Time as Implicit Feedback." Fourteen-week naturalistic study of seven participants; a strong counterexample to universal dwell semantics. [DOI](#)
30. Kim, Y., et al. (2014). "Modeling Dwell Time to Predict Click-Level Satisfaction." Segment-conditioned prediction using editorial labels; not a universal semantic threshold. [DOI](#)
31. Li, J., Huffman, S. B., & Tokuda, A. (2009). "Good Abandonment in Mobile and PC Internet Search." Editorial cross-device and cross-locale study of satisfied no-click possibilities. [Google Research](#)
32. Diriye, A., et al. (2012). "Leaving So Soon?" Retrospective and in-situ prompting study showing multiple abandonment rationales and substantial nonresponse. [Microsoft PDF](#)
33. Ageev, M., et al. (2011). "Find It If You Can." Known-answer task study separating several forms of success and showing an inferential ceiling in natural logs. [DOI](#)
34. Jones, R., & Klinkner, K. L. (2008). "Beyond the Session Timeout." Manual hierarchical annotation showing interleaved and resumed goals in one historical engine. [Author PDF](#)

Evidence methods and error

35. Parry, D. A., et al. (2021). "Discrepancies Between Logged and Self-Reported Digital Media Use." Preregistered systematic review and meta-analysis of 106 effect sizes. [DOI](#)

36. **Sen, I., et al. (2021).** "A Total Error Framework for Digital Traces of Human Behavior." Framework separating representation and measurement errors. [DOI](#)
37. **Bosch, O. J., et al. (2025; online 2024).** "Tracking Undercoverage in Web Tracking Data." Three-country panel evidence and simulations of bias from incomplete browser/device capture. [DOI](#)
38. **Kelly, D. (2009).** Methods for Evaluating Interactive Information Retrieval Systems with Users. Authoritative methodological monograph covering designs, sampling and measures. [Publisher PDF](#)
39. **Connaway, L. S., et al. (2022).** "Speaking on the Record." Study combining discovery-system logs with personalised event-cued interviews. [DOI](#)
40. **Borlund, P., & Schneider, J. W. (2010).** "Reconsideration of the Simulated Work Task Situation." Review of realism and participant relevance in interactive-retrieval experiments. [DOI](#)