ABSTRACT
Lack of theoretical foundation in the design and development of interactive information retrieval (IIR) systems has been the major limitation in the field. Application of information seeking theories and models has also been limited in IIR. This paper is concerned with an approach to facilitate theoretical development in IIR based on cognitive effects, which can be considered in the evaluation of information seeking and retrieval (IS&R) research. In particular, I argue that looking at an “effect” in information seeking might facilitate our communication on IS&R phenomena and theory building in IIR.

1. INTRODUCTION
This paper is concerned with theoretical development in interactive information retrieval (IIR). By theoretical, I mean the use of scientific tools such as framework, theory, law, principle, model, and hypothesis [11]. By development, I mean to continuously build research upon existing research using the scientific tools. We are all aware that theory is the foundation of science. Shoemaker, et al. [17] state that a scientific theory is useful for summarising knowledge, developing practical applications, and guiding research. Nevertheless, we often focus on solving a problem at hand practically by developing systems, interfaces, or algorithms, rather than theoretically. Perhaps, the educational systems tend to an unbalanced weight on teaching practical solutions (e.g., how to build a program) and on teaching theoretical solutions (e.g., how to build and use a theory). Besides, theory building is a difficult job.

The design and development of interactive information retrieval systems have also been criticised as an ad-hoc process and lacking theoretical underpinning [6]. To address this issue, this paper explores ways to facilitate theoretical development in IIR in the context of Information Seeking and Retrieval (IS&R) research. The goals of IS&R research include 1) theoretical understanding of IS&R, 2) empirical explanation of IS&R, and 3) supporting the design and development of information systems [20, 16]. However, my survey of existing studies was originally motivated by the search for a theoretical framework for the design of search interfaces, and thus, the discussions in this paper are sometimes limited to this application, although often they have implications for a wide audience. The range of variables considered in IIR has expanded significantly due to the recent advance of IR in Context. Therefore, the role of theory has become essential for principled advance in this area.

The rest of paper is structured as follows. I first discuss the gap between information seeking models and information retrieval models, and how researchers have bridged it. Then I discuss cognitive effects often observed in Psychology as a potential building block of theoretical development. Finally, I discuss the implications of cognitive effects on the evaluation of search interfaces and other IIR techniques.

It should be noted that this paper does not provide detailed discussions on the definitions of all scientific tools for theoretical development. Interested readers might refer to an excellent paper by Järvelin [11] on conceptual tools in IS&R, and a book by Shoemaker, et al. [17] on theory building.

2. THE GAP
The designers and developers of interactive systems often seek theory to justify the design of (part of) a system and interface. One of the areas they look to is Information Seeking (IS), which studies the use and preference of information sources and channels, or more generally, information seeking behaviour. However, there are certain differences between IS and IIR which make this collaboration difficult. Järvelin and Vakkari noted that “by comparing the typical dependent variables in IR and IS studies it is easy to see that these two fields are interested in difference phenomena. The former is focused on the precision and recall of retrieval methods used by humans in interactive IR. ... The latter is focused on the preferences and use of channels” [20, p.124]. “If dependent and independent variables and their relationships do not overlap in the studies, it is understandable if representatives of both fields do not consider results and ideas from the other field useful” [20, p.124].

Another cause of the gap appears to be the characteristics of theoretical models used in the two fields. In IR, the models tend to be formal (e.g., Language Model), while they tend to be more conceptual in IS [24]. Ingwersen and Järvelin
A model’s ability to generate testable hypotheses is strongly related to predictive power of theory. Rutherford and Ahlgren [15] emphasised the importance of predictive power as follows. “The essence of science is validation by observation. But it is not enough for scientific theories to fit only the observations that are already known. Theories should also fit additional observations that were not used in formulating the theories in the first place; that is, theories should have predictive power. Demonstrating the predictive power of a theory does not necessarily require the prediction of events in the future. The predictions may be about evidence from the past that has not yet been found or studied.” It is the predictive power that facilitates the design and development of practical applications based on a theory. However, in social science, the predictive power is often ignored and explanatory power is the focus of a model [17]. As a consequence, it is difficult to derive testable hypothesis from IS models and to base the design of search systems on them.

However, this trend does not seem to be unique to IS. Gibbs [7] noted that “throughout the field’s history, sociologists have displayed an astonishing tolerance of untestable theories. Indeed, given the veneration in sociology of grand theories and the common indifference of sociologists to the few testable theories in their field, sociologies clearly pay little attention to testability when assessing theories.” The limitation of summary models were echoed by Wilson [24] noting that “it does little more than provide a map of the area and draw attention to gaps in research: it provides no suggestion of causative factors in information behaviour and, consequently, it does not directly suggest hypotheses to be tested.”

There is limited work which has attempted to overcome this gap between the two fields (e.g., [5, 12, 19]). For example, Fidel and Pejtersen demonstrated that Cognitive Work Analysis provided one approach to make studies in information seeking behaviour relevant to system design. A disadvantage of their approach is that it involved an extensive field study involved in the process which may not be available to us, and it is not easily repeatable. Vakkari [19] derived hypotheses from Kuhlthau’s Information Search Process (ISP) model [13], and extended it to IIR context. For example, search tactics in ISP model was simply browsing or querying. Vakkari’s extension contained more detailed query options. This shows that it is possible to bridge the gap between these two fields using scientific tools. However, as emphasised in Järvelin and Vakkari [20], this is partly due to the exceptionally good predictive power of the ISP model. Again, this reinforces the importance of predictive power in theoretical development.

3. COGNITIVE EFFECTS

How can we gain predictive power in the research? A way in which an understanding of human behaviour is represented in Psychology offers us one approach to solve this problem. This is often called an effect. An effect is a result of a cause (i.e., causality). A cognitive effect is a form of change in our cognitive state or process such as perception, learning, problem solving, memory, attention, and language caused by an event. An example of effects which is relevant to IIR research is the novelty effect [4]. The novelty effect refers to our tendency of temporally liking new technologies which diminishes as time goes by. Participants of user studies might prefer an experimental system over a baseline system due to this effect. A large number of effects has been observed and, to a different degree, studied by the researchers in Cognitive and Behavioural Science (See the List of effects section, List of cognitive biases, and List of memory biases section in Wikipedia).

Effects have interesting traits as a conceptual tool of research. First, they can be a smaller unit than a model or theory. An effect might occur in a part of the model. This means that we can present and discuss effects more easily than models or theories. Second, they tend to have predictive power. Effects help us to predict what will happen based on a condition or function (e.g., when a new technology is introduced, people will initially like it). As we have discussed earlier, this allows us to formulate hypotheses to investigate this effect. Third, they do not always have a clearly understood cause, which is the principle of causality. This may sound contradictory to the previous point. However, the condition might represent a situation where an effect occurs, but it does not necessarily explain why it occurs. This means that we can present and discuss an effect without complete understanding. Fourth, effects are not necessarily about new ideas. It can be a conceptualisation of an effect we are empirically familiar with. An example is picture superiority effect, which essentially says concepts are more likely to be remembered if they are presented as pictures rather than as words. This simple effect, nevertheless, provides a theoretical justification for using a thumbnail of webpages as a form of bookmarking in search systems. Finally, some effects are a compound of other effects. For example, serial position effect shows that when people remember a sequence of items, they can recall the start and end items better than the middle items. This effect is a result of the primacy effect (for the first items) and the recency effect (for the last items).

An excellent example of cognitive effect in IS&R is the work on task complexity by Byström and Järvelin [3]. The overall effect of task complexity in their study can be defined as the changes of the needs of information caused the people to.
ceved complexity of the task at hand. More specifically, it models that as task complexity increases 1) the complexity of information needed increases; 2) the needs of domain information and problem solving information increase; 3) the share of general-purpose sources increases, and that of problem and fact-oriented sources decreases; 4) the success of information seeking decreases; 5) the internality of channels decreases; and 6) the number of sources increases [3]. As can be seen, the effects help us predict a consequence for a given change. This helps us to design a responsive support for a given situation. The effects are also testable hypotheses [21]. Another task complexity effect was people's searching behaviour. For example, Machionini [14] observed that people are more likely to perform keyword search in low complexity tasks while they are more likely to perform browsing in high complexity tasks. Some of the accounts made by Byström and Järvelin help us understand the cause of this effect. Now we can see that a potential theory of task complexity in IS&R has been developing as a result of accumulated studies that look at effects and their causes.

Van Rijsbergen's clustering hypothesis [22] is another example. It hypothesises that similar documents are likely to be relevant for the same query. This hypothesis has been examined by numerous studies (e.g., [18]) and thus has become one of the most frequently used techniques in information systems. Arguably, it was the form of an effect, document similarity to topical relevancy, represented in the hypothesis that allowed researchers to use it as a building block of more complex applications.

Unfortunately, we do not have many cases like task complexity or the clustering hypothesis in IS&R. The next section discusses the implications of cognitive effects on IS&R research.

4. IMPLICATIONS ON IS&R EVALUATION

There are several directions of research based on cognitive effects. First, we can investigate existing cognitive effects found in Psychology in the IS&R context. Novelty effect, for example, is often used to describe a short-time positive effect on people's task performance or perception caused by a new technology. A new technology tends to increase a level of attention, which results in increased efforts or persistence, which in turn yields an achievement gain [4]. However, when they increase their familiarity with the technology this positive effect is known to disappear. Therefore, when participants of a user study are asked for their preference towards a baseline system (with low novelty) or an experimental system (with high novelty), a higher level of preference on the experimental system might be a result of its higher novelty rather than increased level of effectiveness, efficiency, or usability.

However, the increased level of effort and persistency caused by novelty appear to have more implications on IS&R. Recently, there have been several studies that tested the effectiveness of search interface that offer a form of grouping function that allows users to organise search results [23]. This type of interfaces is often designed to support exploratory tasks where information needs and task goals are ill-defined, and thus, iterative searching and browsing are required. One of the common results in such studies is that the number of interactions with system increased in experimental systems (with grouping function) when compared to baseline systems (without grouping). As discussed above, there is technological novelty in experimental systems. However, there also appears to be topical novelty. In other words, participants were finding new relevant information or new aspects/facets/instances in a topic with experimental systems, which could encourage them to make more effort and to be more persistent with artificial tasks. Therefore, continuously providing topical novelty might be a key factor to facilitate exploratory tasks, and thus, the design of exploratory search interfaces.

A second direction is to revisit existing IS&R findings to derive cognitive effects. The task complexity effect was such as example. Belkin, et al. [1] found that people's search queries can be lengthened when a short text "Information problem description (the more you say, the better the results are likely to be)" was added to the bottom of the query box. Since there is a higher degree of ambiguity in short queries, longer queries tend to perform better in search. This might be called the query box effect. Another example is click-through behaviour from query log analysis. It often shows a high click rate on the top ranked documents (e.g., 1st, 2nd, 3rd), but the URLs shown at the bottom of the result page (10th) are sometimes clicked more frequently than the middle (5th-9th). This is often explained by a scrolling action or paging action. When searchers try to go to the next page of search result, they saw the 10th result which was hidden before scrolling. This might be called the page navigation effect. How can we prevent this in terms of search interface design? Or is it worth placing a higher score result in the 10th place than 5-9th?

A third direction is to derive an effect by looking at the cases where proposed techniques did not work. This is a type of failure analysis. Although the designer of the system or search interface often assumes rational behaviour of the users, cognitive bias and irrational behaviour are common and persistent in human activities. By examining the cause of unsuccessful performance, one may hypothesise a potential effect that explains the situation. This can be a more constructive way of reporting our negative result rather than insignificant performance differences.

Readers may find the use of the label "effect" inappropriate and may prefer to use "hypothesis". As I discussed above, both have common traits, although a hypothesis is not necessarily about a casualty. The anomalous state of knowledge (ASK) hypothesis [2] is such an example. The ASK is about the nature of information needs. "Bias" and "fallacy" are also frequently used for a negative effect in Psychology. However, a more prominent assertion I would like to make in this paper was to encourage researchers to observe and formulate an effect as a part of evaluation and present it as a theoretical contribution of the research. The observed effect might be weak or sparse. You may not have a complete explanation, or later we might find that it was a false effect. However, at least this enables other researchers to use it as a building block of more complex applications.
Measures used in laboratory-based experiments of IIR studies have been predominantly retrieval effectiveness (e.g., precision, recall, cumulative gain), interaction efforts or efficiency (e.g., task completion time, number of queries, frequency of browsing), and subjective assessments (e.g., task perceptions, system perceptions). This has resulted in a number of descriptive studies without much theoretical development. Shoemaker, et al. [17] noted “scholarly journals offer a wide variety of descriptive studies. ... These kinds of atheoretical research areas have their uses but they tend to produce isolated studies that do not move our knowledge forward on important questions in the field. ... such studies become like bricks lying around the brickyard rather than bricks that are used to build a wall” (p.168).

A strong tradition of measuring improvement in IR might have been preventing us from deriving potential effects from studies. Finding a cognitive effect in an evaluation can be a starting point of theoretical development as illustrated in this paper. The form of effects appears to be worth considering to represent our understanding of IS&R phenomena and to facilitate our communication about the phenomena.

5. SUMMARY
As discussed in the Introduction, we often focus on developing practical solutions to a problem, putting theoretical development at a lower priority in research. Also, theory building can be a difficult and abstract task. Exploring an effect in our studies might offer us a light-weight starting point for theoretical development which might eventually lead to an interesting theory. As Interactive IR is sometimes called Cognitive-oriented IR [10], the field has considered cognitive aspects in information seeking since its early days (e.g., [2, 9]). This paper discussed another aspect, a much smaller aspect, to contribute to the development of cognitive aspects in Information Seeking and Retrieval.

6. ACKNOWLEDGEMENT
The author thanks Robert Villa for useful discussions during the preparation of this paper. The author is also grateful for the constructive comments from anonymous reviewers.

7. REFERENCES