HEC MONTRÉAL

# À travers les yeux de l'expertise : Comparaison de l'attention visuelle des analystes d'affaires et des novices en modélisation conceptuelle

Par

Karl-David Boutin

Sciences de la gestion
(Technologies de l'information)

Sous la direction de
Pierre-Majorique Léger, Ph.D.

*Mémoire présenté en vue de l'obtention*
*du grade de maîtrise ès sciences*
*(M.Sc.)*

Décembre 2018

# Sommaire


La modélisation conceptuelle étant une activité majeure dans le domaine des Systèmes d'Information (SI), il est primordial de comprendre comment, au niveau cognitif, les modeleurs comprennent et conçoivent des modèles de processus à l'aide de notations visuelles, afin d'affiner les notations et d'améliorer la formation des futurs analystes d'affaires. Ce mémoire par article étudie donc les heuristiques des analystes d'affaires lors de tâches de détection d'erreurs dans des modèles conceptuels en BPMN.

Pour ce faire, une étude en laboratoire a été faite en comparant l'attention visuelle, à l'aide de l'oculométrie, de 15 analystes d'affaires avec celle de 15 novices, lorsqu'ils devaient identifier et diagnostiquer des erreurs dans 75 modèles conceptuels. Chaque modèle avait soit une erreur sémantique, soit une erreur syntaxique ou aucune erreur.

Nos résultats démontrent que la différence entre les experts et les novices est étonnamment modeste, ce qui offre une nouvelle perspective sur la richesse et l'importance de l'aspect cognitif de l'expertise en modélisation conceptuelle. D'ailleurs, nous soulignons la nécessité d'approfondir la littérature sur l'essence de l'expertise en contexte de modélisation de processus d'affaires et sur la nature des différences entre analystes d'affaires et novices. De plus, les résultats de ce mémoire contribuent à approfondir la compréhension des différences entre les erreurs sémantiques et syntaxiques, ce qui permet de poser des lignes directrices en ce qui concerne l'amélioration des techniques et curriculums de formation pour les nouveaux analystes d'affaires.



**Mots clés :** Modélisation conceptuelle, Attention visuelle, Expertise, Analyste d'affaires, Détection d'erreur, Oculométrie.

# Table des matières

# Liste des tableaux et figures

## Liste des figures

**Premier article**



**Deuxième article**



## Liste des tableaux



**Deuxième article**

# Avant-propos

Ce mémoire, rédigé sous la forme de deux articles, a été approuvé par la direction administrative du programme de la Maîtrise ès sciences en gestion et le consentement des coauteurs des deux articles composant le mémoire a été obtenu.

L'approbation du comité d'éthique de la recherche (CER) de HEC Montréal a été reçue pour cette étude en mars 2017.

Le premier article a été soumis et accepté à la conférence Vienna Retreat on NeuroIS 2018, à Vienne, Autriche. L'article est paru dans dans *Information Systems and Neuroscience - NeuroIS Retreat 2018*, dans la série de *Lecture Notes in Information Systems and Organisation,* publié par Springer (Boutin et coll., 2019).

Le deuxième article actuellement en préparation pour soumission à la revue *Journal of Management Information Systems* (JMIS).

Les articles ont été ajoutés au mémoire avec le consentement signé des coauteurs.

# Remerciements

Je voudrais prendre le temps de mentionner et remercier toutes les personnes qui m’ont soutenu lors de mon parcours à la maîtrise et la rédaction de mon mémoire.

Tout d’abord, je tiens à remercier mon directeur de recherche Pierre-Majorique Léger pour son support, ayant pris le temps de me guider dans le monde de la recherche qui m’était encore inconnu. Son aide et sa confiance m’ont permis de vivre des expériences à la fois enrichissantes et excitantes et à me partager son intérêt pour la science. Mon expérience au Tech3Lab fut exceptionnelle et formatrice pour mon parcours professionnel.

Je veux aussi remercier mes co-chercheurs Alan R. Hevner et Christopher J. Davis, pour leurs aide, patience et amabilité. Ce fut un grand honneur et plaisir de travailler avec eux sur ce projet.

Je tiens également à remercier tous mes collègues du Tech3Lab et d’ERPsim, des gens exceptionnels qui m’ont aidé dans mes difficultés et motivé quand le moral était bas. Un merci particulier à Émilie, Michael et Hubert-David, sans qui la maîtrise n’aurait pas été qu’une autre année d’école sans goût et plaisirs. Merci aussi à tout le personnel du Tech3lab ayant aidé au bon déroulement de mon projet.

Un grand merci aussi à toi Félix ! Tes conseils et ta fameuse sagesse ont été inestimables dans mes moments d’hésitation ou de confusion. Je crois que tu ne comprendras jamais à quel point tu m’as aidé au cours de la dernière année, et je t’en remercie du fond du cœur.

Finalement, un merci à ma famille, Cassandra, Lyne, et Marc-André, et à tous mes ami(e)s, plus particulièrement Maxime et Réda, pour m’avoir enduré durant cette dernière année. Je ne sais pas ce que j’aurais fait sans votre support et vos encouragements. Merci infiniment

# Chapitre 1 : Introduction

## 1.1 Mise en contexte

Depuis la révolution numérique, débutée vers la fin des années 1950, les technologies de l'information (TI) n'ont cessé de jouer un rôle grandissant dans les organisations. Même aujourd'hui, quelques décennies plus tard, les TI restent un moteur de création de valeur, d'innovation et de différenciation pour les entreprises dans tous les domaines et marchés (Kark et coll., 2017). En 2018, le budget moyen mondial des TI a augmenté de 6.2% par rapport à 2017 et est prévu d'augmenter d'environ 3% par année, pour les 4 prochaines années (Gartner, 2018). Cette importance croissante des TI a aussi amené un changement de paradigme important sur l'architecture des entreprises et la gestion de ses processus.

Ce changement de paradigme, qui prône l'intégration des différents unités et domaines fonctionnels de l'entreprise a, quant à lui, permis le développement de la gestion des processus d'affaires (GPA). Cette discipline, qui reste toujours une priorité des entreprises aujourd'hui, permet de cerner les processus d'affaires des entreprises afin de les optimiser par, notamment, la réingénierie de processus (Recker et coll., 2009). Plusieurs approches à la gestion des processus d'affaires ont été développées, du système de production Toyota (TPS) à l'approche Six Sigma, ce qui a grandement stimulé la recherche sur cette discipline et les différents outils facilitant ces activités (Jeston et Nelis, 2010 ; Recker et coll., 2009).

La croissance des investissements en GPA et, par le fait même, de la recherche dans ce domaine a amené le développement de la modélisation des processus d'affaires et la multiplication des différentes techniques et formalismes sous-jacents (Recker et coll., 2009). À travers les années, la modélisation des processus d'affaires est devenue une compétence clé des analystes d'affaires, permettant la communication et le partage des connaissances, la conception et l'amélioration des processus et la redéfinition des décisions dans les organisations (Recker et coll., 2012 ; Becker et coll., 2004 ; Indulska et coll., 2009). De plus, l'utilisation de notation visuelle, telle que l'UML, le BPMN ou l'EPC, permet de créer des modèles compréhensibles par les parties prenantes de différentes unités organisationnelles, ou même différentes organisations (Davis et coll., 2018).

Toutefois, la prolifération des techniques de modélisation entraîne des complications importantes au niveau de la formation des analystes d'affaires et du développement de logiciels supportant la gestion des processus d'affaires (Recker et Dreiling, 2007 ; Aguilar-Savén, 2004). En effet, parmi la panoplie de choix disponibles, il est maintenant nécessaire pour les établissements scolaires, développeurs de logiciels et entreprises de décider quel(s) formalisme(s) à utiliser et à enseigner aux futurs analystes d'affaires (Recker et Dreiling, 2007).

De nombreuses études ont alors examiné la modélisation conceptuelle des processus d'affaires sous un nombre impressionnant d'angles différents, par exemple, en étudiant l'effet de différentes notations et des connaissances antérieures liées au domaine ou à la modélisation sur la compréhension de modèles conceptuels (Yusuf et coll., 2007 ; Ottensooser et coll., 2012 ; Rodrigues et coll., 2015 ; Recker and Dreiling, 2007). De plus, un éventail de recherches a étudié l'effet des caractéristiques des modèles (tels que la disposition du modèle, les étiquettes et conventions de nommage des activités ou la grosseur des modèles) sur la compréhension de ceux-ci (Figl, 2017). Cependant, tout en évaluant les notations ou leurs caractéristiques, la composante syntaxique est rarement discutée, ce qui représente une lacune importante dans la littérature, car les différences syntaxiques entre les notations sont aussi importantes, sinon plus importantes, que les variances sémantiques (Moody, 2009).

Finalement, bien que la littérature sur la modélisation des processus d'affaires soit bien étoffée, l'utilisation de notation visuelle pour modéliser graphiquement des processus reste, en pratique, très complexe ; il est, en effet, pratiquement impossible de concevoir un modèle sans ambiguïtés (Figl, 2017). De plus, les intérêts de la communauté académique ne sont pas toujours alignés avec les besoins des analystes d'affaires (Indulska et coll., 2009 ; Jabbari et coll., 2017 ; Wand et Weber, 2002). Effectivement, certains domaines d'intérêts des analystes d'affaires, tels que l'expertise et la formation, sont peu étudiés par la communauté académique (Indulska et coll., 2009). Ce manque de connaissances sur la nature de l'expertise en modélisation de processus d'affaires dans la littérature est un obstacle important au développement et à l'amélioration des programmes de formation (Davis et coll., 2018).

## 1.2 Objectifs de l'étude et questions de recherche

La modélisation conceptuelle des processus d'affaires, bien qu'ayant toujours été une activité centrale dans le domaine des systèmes d'information (SI) utilisés principalement en tant que moyen de communication ou de partage des connaissances et comme outil dans l'amélioration des processus d'affaires dans les organisations, fait l'objet de plus en plus d'études (Recker et coll., 2012 ; Becker et coll., 2004 ; Indulska et coll., 2009). Toutefois, certains concepts liés à la modélisation conceptuelle restent encore nébuleux et posent de grandes difficultés aux analystes d'affaires (Indulska et coll., 2009 ; Eikebrokk et coll., 2008). Ces lacunes dans la littérature offrent des opportunités importantes pour contribuer à la connaissance et identifier des facteurs pouvant aider à affiner les notations et améliorer la formation des futurs analystes d'affaires.

Ce mémoire se penche donc sur les caractéristiques de l'attention visuelle dans une tâche quotidienne des analystes d'affaires, soit l'identification d'anomalie dans des modèles conceptuels.

Le premier article cherche à développer notre compréhension de l'attention visuelle en modélisation conceptuelle et nous familiariser avec l'utilisation de l'oculométrie dans ce contexte. L'objectif du premier article est donc d'explorer les différences dans les caractéristiques de l'attention visuelle entre les diagnostics réussis et les échecs dans une tâche de détection d'erreurs en modélisation conceptuelle.

En utilisant les résultats et observations du premier article, la deuxième étude cherche à comparer les techniques des novices et experts modeleurs lors de la même tâche de détection d'erreurs, afin d'approfondir la compréhension de la dichotomie expert-novice en modélisation conceptuelle. De plus, en étudiant les différences entre la détection d'erreurs sémantiques et syntactiques, nous cherchons à identifier les éléments les plus complexes des modèles de processus d'affaires.

En comparant les novices avec les experts, ce mémoire tente d'identifier les « meilleures pratiques » et les éléments les plus complexes de la modélisation conceptuelle, afin de pouvoir faire des recommandations sur le développement et l'amélioration des programmes de formation des analystes d'affaires. Ce mémoire tente donc de répondre à la question de recherche suivante :

*Quelles sont les différences relatives à l'attention visuelle entre les analystes d'affaires experts et les novices, lors d'une tâche de détection d'erreurs dans des modèles conceptuels ?*

## 1.3 Structure du mémoire

Ce mémoire par article comporte deux articles portant sur la même expérience. Le premier article est sous la forme d'acte de conférence. Réalisé à la fin de la première collecte de données, cet article rapporte les résultats préliminaires sur un échantillon de 18 participants et permet de poser les bases conceptuelles pour le second. Il permet d'approfondir notre compréhension des caractéristiques de l'attention visuelle sous-jacentes aux diagnostics exacts. Le deuxième article, quant à lui, est plus volumineux puisque destiné à un journal. Il a été écrit après la 2e collecte de données qui était une réplication de la première, et rapporte les résultats complets pour l'ensemble des 30 participants. Il compare aussi les caractéristiques de l'attention visuelle liées aux bons et mauvais diagnostics, mais inclut une comparaison entre novices et experts afin d'identifier les techniques et faiblesses des analystes d'affaires. Pour conclure, nous reviendrons sur les principaux résultats, contributions et limitations des deux articles.

## 1.4 Contributions potentielles

D'un point de vue théorique, ce mémoire contribue à approfondir la littérature sur la dichotomie novice-expert, en cherchant à comprendre la nature de l'expertise, ainsi que les forces et faiblesses des experts en modélisation conceptuelle. De plus, en comparant les processus d'identification d'erreur sémantique et syntaxique, nous répondons au manque d'étude sur l'effet des règles syntaxiques sur la compréhension des modèles (Moody, 2009). Finalement, en étudiant les techniques utilisées par les experts, nous cherchons à identifier les « meilleures pratiques » dans l'analyse de modèles conceptuels et les limites des notations visuelles.

D'un angle pratique, étudier les techniques de lecture et d'analyse de modèles conceptuels des experts nous permettra d'offrir des recommandations claires sur le développement et l'amélioration des curriculums et méthodes de formation des futurs analystes d'affaires. De plus, une connaissance approfondie des différences entre les caractéristiques de l'attention visuelle liées à la détection d'erreurs sémantiques et syntaxiques nous permettra de poser des recommandations afin d'orienter l'évolution de la notation BPMN, ainsi que les autres notations visuelles.

## 1.5 Information sur les articles

### 1.5.1 Article 1

Le premier article a été soumis et accepté à la conférence scientifique Vienna Retreat on NeuroIS 2018, à Vienne en Autriche, une association ayant pour intérêt l'utilisation d'outils et de méthodes liés à la neuroscience pour mieux comprendre l'utilisation et l'impact des technologies de l'information. Les premières phases du projet ont débuté en hiver 2018 par l'étudiant de ce mémoire sous une bourse de recherche du CRSNG. La collecte de données a été effectuée en mars 2018 par l'étudiant et les résultats ont été présentés à la conférence en juin 2018 (Boutin et coll., 2019).

#### 1.5.1.1 Résumé de l'article 1

Nous utilisons l'oculométrie pour mieux comprendre les caractéristiques de l'attention visuelle spécifiques à la détection d'erreurs réussie dans des modèles conceptuels. Nous testons nos prédictions basées sur des études antérieures sur l'attention visuelle dans des tâches de détection d'erreurs, ou des études comparant des experts et des novices dans diverses tâches, dans une expérience contrôlée où les participants sont chargés de détecter et de diagnostiquer les erreurs dans 75 modèles en BPMN. Les résultats suggèrent que les diagnostics d'erreurs réussis sont liés à un temps de vue total des stimuli plus court et à une durée de fixation plus courte, avec une différence significative entre les erreurs sémantiques et syntaxiques.

### 1.5.2 Article 2

Le deuxième est en cours de finalisation pour être soumis au journal scientifique *Journal of Management Information Systems* (JMIS). Une deuxième collecte de données a été complétée par l'étudiant de ce mémoire en mai 2018. L'article tient compte des données utilisées dans le premier article en plus des données récoltées lors de la deuxième collecte. Une version préliminaire de cet article est présentée dans ce mémoire.

#### 1.5.2.1 Résumé de l'article 2

Nous étudions les caractéristiques de l'attention visuelle propres aux experts et aux novices dans le cadre d'une tâche de détection d'erreur sémantique et syntaxique à travers 75 modèles conceptuels écrits en BPMN. Nos résultats suggèrent que les experts ont plus tendance à diagnostiquer de fausses anomalies dans des modèles n'ayant aucune erreur que les novices. Toutefois, les experts les plus performants distinguent mieux les stimuli sans erreur que les autres participants. Nos résultats concernant la légère différence entre experts et novices mettent en évidence la non-polarité du concept d'expertise et la nécessité d'étudier davantage la manière dont les analystes d'affaires analysent les modèles conceptuels.

## 1.6 Contributions personnelles

Tableau 1 : Contributions et responsabilités dans la rédaction des articles

| Étape | Contribution et tâches effectuées |
|---|---|
| Définition des requis | Définition de la question de recherche et la problématique – 100 %<br>● Problématique existante à l'initiation du projet<br>● L'équipe a contribué à la définition des questions de recherche et l'approche à adopter |
| Revue de la littérature | Effectuer la revue de littérature pour déterminer les construits testés dans le domaine de l'expertise et la modélisation conceptuelle – 100 %<br>Définir et proposer les outils de mesure à utiliser pour tester les construits – 80 %<br>● Le reste de l'équipe de recherche s'est assuré que les outils de mesure sélectionnés permettaient de tester les construits choisis |
| Stimuli | Conception des stimuli pour l'expérience – 100 %<br><br>Modification et correction des stimuli après les prétests - 90 % |

| | |
|---|---|
| | <ul><li>Soutenue par mon co-chercheur Christopher J. Davis</li></ul> |
| Conception du design expérimental | Création des formulaires nécessaires pour la demande au CER – 100 %<br><br>Concevoir le protocole de l'expérience – 100 %<br><br>Création du matériel de formation pour l'expérience - 100 %<br><br>Préparation de la salle de collecte - 90 %<ul><li>L'équipe a ajusté les instruments oculométriques lors de la première collecte</li></ul> |
| Recrutement des participants | Élaborer le formulaire de recrutement – 100 %<br><br>Sollicitation et recrutement des participants – 100 %<br><br>Gestion des horaires de participation - 100 %<br><br>Responsable des compensations – 100 %<br><br>Concevoir le cartable d'expérience pour le suivi des participants – 100 % |
| Prétests et collecte | Responsable des prétests - 100 %<br><br>Responsable des collectes de données - 100 % |
| Extraction et transformation des données | Déterminer les mesures à extraire afin de faire les analyses statistiques nécessaires - 100 %<br><br>Extraction et nettoyage des données oculométriques et provenant des questionnaires - 100 % |
| Extraction et mise en forme des données | Extraction et mise en forme des données oculométriques – 100 % |

| | |
|---|---|
| Analyse des données | Analyse des données oculométriques – 100 %<br>● Création de 75 aires d'intérêt<br><br>Analyses statistiques du mémoire – 80 %<br>● Aide sur Stata/MP 15.1 pour les analyses par le statisticien de la Chaire de recherche industrielle CRSNG-Prompt en expérience utilisateur |
| Rédaction | Contribution dans l'écriture des articles – 100 %<br>● Les autres auteurs m'ont donné des commentaires tout au long de la rédaction afin d'améliorer la qualité des articles. |

# Chapitre 2: Premier Article

## Attentional Characteristics of Anomaly Detection in Conceptual Modeling

Karl-David Boutin[1], Pierre-Majorique Léger[1], Christopher J. Davis[2], Alan R. Hevner[3], and Élise Labonté-LeMoyne[1]

[1] HEC Montréal, Québec, Canada
{karl-david.boutin, pierre-majorique.leger, elise.labonte-lemoyne}@hec.ca
[2]University of South Florida, Saint Petersburg, FL, USA
davisc@mail.usf.edu
[3]University of South Florida, Tampa, FL, USA
ahevner@usf.edu

**Abstract:**

We use eye tracking to better understand the attentional characteristics specific to successful error detection in conceptual models. This phase of our multi-step research project describes the visual comportments associated with successful semantic and syntactic error identification and diagnosis. We test our predictions, based on prior studies on visual attention in an error detection task, or studies comparing experts and non-experts in diverse tasks, in a controlled experiment where participants are tasked with detecting and diagnosing errors in 75 BPMN® models. The results suggest that successful error diagnostics are linked with shorter total view time and shorter fixation duration, with a significant difference between semantic and syntactic errors. By identifying the visual attention differences and tendencies associated with successful detection tasks and the investigation of semantic and syntactic errors, we highlight the non-polarity of the 'scale' of expertise and allow clear recommendations for curriculum development and training methods.

**Keywords:** eye tracking · conceptual modeling · attentional characteristics

## 1. Introduction

Business process modeling has become a central activity in IS practice (Recker and Dreiling, 2007). Conceptual models facilitate communication about business domains and their processes (Parsons and Cole, 2005; Gemino and Wand, 2000; Gemino and Wand, 2004). Such models have become a primary medium used in design activities. This phase of our research strives to deepen understanding of visual attention during error detection tasks (Gegenfurtner et al., 2011). While researchers have explored the variations between novice and experienced modelers (Shanks, 1997), the differences in the visual attention between successful and unsuccessful error detection tasks in conceptual modeling are yet to be deeply investigated.

For this research, we employ the Business Process Modeling Notation (BPMN®), an international standard for business processes. Its popularity in commercial settings prompted its selection for this phase of our work. Visual notations such as BPMN are composed of visual syntax - symbolic vocabulary and grammar - and visual semantics - elements that give meaning to each symbol and symbol relationship (Davis et al., 2018; Moody, 2009). However, while evaluating notations or their usage, the syntactic component is rarely discussed (Moody, 2009). This presents an opportunity to contribute to the literature by comparing both the semantic and syntactic error identification process. The main objective of this study is to explore the differences in the attentional characteristics between successful and unsuccessful diagnostics in a detection task. We use eye tracking to monitor the visual attention of subjects as they search for and diagnose semantic and syntactic errors in a controlled experiment.

## 2. Prior Research and Hypotheses Development

Studies regarding the difference in the visual attention in an error detection task conclude that, on average, errors are fixated more often and longer than irrelevant information (Van Waes et al., 2009; Henderson and Hollingworth, 1999; Holmqvist et al., 2011), and that a high number of fixations on the stimulus is correlated with an ineffective search (Holmqvist et al., 2011; Goldberg and Kotval, 1999). Furthermore, the longer the participant spends looking for an error, the lower the chances of success (Van Waes et al., 2009), possibly due to too much cognitive resource being drawn away for the encoding of the stimulus. Studies that compare novices and experienced modelers point to attentional characteristics that might be associated with expertise, and thus, generally, with better performance (Recker and Dreiling, 2007; Yusuf et al., 2007).

Several meta-analyses that use eye tracking to compare experts and novices in a range of domains conclude that those classified as experts spend less time looking at stimuli before fixating relevant areas or anomalies (Gegenfurtner et al., 2011; Krupinski, 2000; Reingold and Sheridan, 2011; Sheridan and Reingold, 2014). More efficient scan patterns (Yusuf et al., 2007; Krupinski, 2000; Reingold and Sheridan, 2011) or more detailed and completed schemata (Glaser, 1984; Lurigio and Carroll, 1985) are offered as explanations. Experts also tend to have fewer fixations, suggesting less cognitive effort to decipher and understand the stimuli (Yusuf et al., 2007; Reingold and Sheridan, 2011), and shorter fixation durations (Gegenfurtner et al., 2011), which are also associated with lower cognitive processing effort (Holmqvist et al., 2011). Therefore, we propose three study hypotheses:

*H1 — Successful error detections in conceptual modeling will require less time spent looking at the stimulus than unsuccessful error detections.*

*H2 — Successful error detections in conceptual modeling will require, in total, fewer fixations than unsuccessful error detections, but with a higher proportion of fixations on the error.*

*H3 — Successful error detections in conceptual modeling will require, on average, shorter fixation duration than unsuccessful error detections, but with longer fixation duration on the error.*

# 3. Research Method

Our experiment was conducted on a sample of 18 participants (7 males, 11 females) with different ages and experience. All our participants were offered a $20 gift card as a compensation for their participation. The research was approved by our institution's Research Ethics Board (REB), and each participant signed a consent form.

## 3.1 Research Design and Protocol

In order to confirm our hypotheses, we tasked our participants with identifying and diagnosing errors in conceptual models written in BPMN. Each participant had to inspect five (5) distinct sets of 15 models (for a total of 75 models), where each set represented a business process scenario (e.g. airport check-in process). An example can be seen in Figure 1. Five (5) versions of each scenario were presented as BPMN 'sentences'. These were further manipulated to present three (3) versions: one with no known errors; one with a known semantic error, and one with a known syntactic error (Davis et al., 2018).

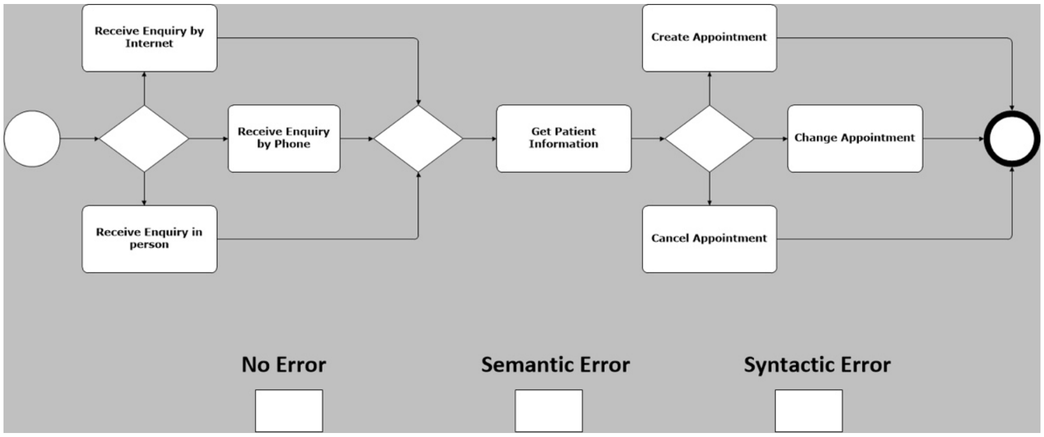



**Fig. 1.** Example of a model with boxes to indicate type of error detected

To mitigate the effect of prior knowledge of the business domains of the models (Gemino and Wand, 2004; Birkmeier et al., 2010), the stimuli were created using simple scenarios, well-known to all potential participants. Furthermore, we limited the range of symbols used and the scope of the 'sentences' to 10-12 elements, favoring numerous accessible models rather than more complex and domain knowledge-dependent stimuli. To train the participants and mitigate the effect of learning through the trial, the experiment started with a short presentation on BPMN (Bavota et al., 2011). The symbols used in the study, as well as the two different types of errors, were shown and explained. The training was concluded with a practice task where participants were shown three (3) models, each one with a different type of error. Just like in the real task, the participant had to pinpoint the error and to diagnose the error type, both by clicking in the corresponding area on the stimuli. The correct location of the error, as well as the right error type, was revealed after each practice model. To avoid bias, the models used in the practice task were not related to the sets of models used later in the experiment, and the conditions were the same as in the experiment (Bavota et al., 2011).

The participants then started the first task with their first set of models. The fifteen models included in the set were shown in random order and without any time limit. After identifying and diagnosing the error in a model, participants had to manually advance to the next model, using the space bar on their keyboard. They then proceed to the next set of models until they completed the five (5) sets.

### 3.2 Apparatus and Measures

Eye tracking (Red 250, SensoMotoric Instruments GmbH, Teltow, Germany) was used to gather the behavioral measures throughout the experiment, at a sampling frequency of 60 Hz. The number of fixations, which is the stabilization of the eye on an object (Yusuf et al., 2007), and their duration were gathered for each area of interest (AOI), as the literature tends to agree that fixation is related to the cognitive processing of visual information (Yusuf et al., 2007; Just and Carpenter, 1976). The fixation duration threshold was set at 200 ms (Holmqvist et al, 2011; Rayner, 1998). The time before the first fixation in an AOI and the total view time of a stimulus were also collected. One (1) to three (3) AOIs were placed on the correct choice of error type and on the actual location of the error(s). For each participant, the eye tracker was calibrated to a maximum average deviation of 0.5 degrees, using a 9-points predefined calibration grid.

## 4. Preliminary Results

We briefly present several preliminary results from our study. Hypothesis 1 states that successful identification and diagnosis of errors in conceptual models will take less time than unsuccessful answers. A linear regression with mixed model and a two-tailed level of significance is performed to compare the Total View Time for each value of the variable Answer (i.e. if the participant successfully diagnosed the error, Answer = 1, if not, Answer = 0). Results suggest that successful detection of error, including models without an error, is linked with lower total time spent on each stimulus ($B$ = -0.3934, p <.0001). Furthermore, successful detection of semantic errors shows a faster time to first fixation on the area of interest (i.e. the zone containing the error) ($B$ = -0.3333, p = .0027). However, there are no significant results linking the detection of syntactic or no errors with the time to first fixation.

Hypothesis 2, which stipulates that successful error detection will be linked with fewer fixations, but with a higher proportion of fixations on the error, is tested using a Poisson regression with mixed model and a two-tailed level of significance of Fixation Count on Answer. A significant relation is found between successfully detecting an error in a model and lower fixation count ($B$ = -0.4402, p <.0001). Moreover, greater proportions of fixation are allocated to the zone containing semantic errors ($B$ = 0.5448, p <.0001) and syntactic errors (*B* = 0.9379, p <.0001). However, while correct diagnosis of semantic errors are linked with a decrease in the number of fixation in the areas of interest (*B* = -0,04956, p = .0897), the opposite is found for the successful detection of syntactic errors, where more fixations on the AOIs are required (*B* = 0,3654, p <.0001).

In order to test Hypothesis 3, we apply a linear regression with mixed model and a two-tailed level of significance of Fixation Durations on Answer, allowing us to identify a significant correlation between successful diagnostics and shorter fixation duration (*B* = -0.373, p <.0001). However, fixations in the AOIs are longer for successful diagnosed semantic errors (*B* = 0.1654, p = .0061) and syntactic errors (*B* = 0.4436, p <.0001), which indicate that the participants, when successfully identifying the errors, spend more time on the errors, but less time on the rest of the stimuli.

## 5. Discussion and Conclusion

Our preliminary results suggest that H1, H2 and H3 are partially supported. Significant links between successfully detecting an error and a lesser amount of time spent on a stimulus (H1), and between an accurate diagnostic and shorter fixation duration are found (H3). Furthermore, H2, which proposed fewer fixations and a greater proportion of fixations on the errors when successfully detecting an error, is supported. However, successful detection of syntactic errors is significantly associated with a greater number of fixations in AOIs, which suggests that the error was detected, but the correction response is inhibited (Van Waes et al., 2009), possibly due to a higher level of complexity in syntactic errors. No significant link between syntactic errors and the number of fixations in the entire stimulus is found. Thus, this study presents evidence that the characteristics of visual attention of experienced modelers, such as lower number and duration of fixations, are generally related with successful error detection. On the other hand, attributes normally associated with novices, such as higher time spent on a stimulus or higher fixation duration, lead to unsuccessful error detection.

Our research so far offers both theoretical and practical contributions. The differences between the process and repertoire of attentional characteristics in the detection of semantic errors and syntactic errors reinforce Moody's (Moody, 2009) propositions about their complex interdependence. Syntactic errors require more attentional fixation than semantic errors. This finding runs contrary to our expectations and highlights the need for further studies to more fully articulate the differences and the metrics that might be used to measure them. The next phase of our research will address this challenge. From a practical standpoint, deeper insights into differences between the attentional characteristics will offer guidance to the evolution of BPMN and other notations and recommendations for curriculum development and training methods.

Limitations of this exploratory phase of our study center on the models used as stimuli for our experiment. While we tried to minimize the impact of domain-specific knowledge by using processes well-known to all potential participants, it is virtually impossible to negate the influence of the variation of domain familiarity between participants. Another limitation can be found in our sample. A bigger and more equally distributed sample will allow more complex statistical analyses and provide more significant findings. Finally, no measure was taken to evaluate the 'stopping rule', which is when a subject decides to terminate his information search because he judges that he has enough information to complete his task (Browne and Pitts, 2004; Nickles et al., 1995). The next step of our research should evaluate this concept in order to better understand our eye tracking data, especially the measures linked to the view time.

## References

Bavota, G., Gravino, C., Oliveto, R., De Lucia, A., Tortora, G., Genero, M., Cruz-Lemus, J.A.: Identifying the weaknesses of UML class diagrams during data model comprehension. In: International Conference on Model Driven Engineering Languages and Systems, pp. 168-182. Springer (2011)

Birkmeier, D., Kloeckner, S., Overhage, S.: An Empirical Comparison of the Usability of BPMN and UML Activity Diagrams for Business Users. In: ECIS, p. 2 (2010)

Browne, G.J., Pitts, M.G.: Stopping rule use during information search in design problems. Organizational Behavior and Human Decision Processes 95(2), 208-224 (2004)

Davis, C.J., Hevner, A.R., Labonte-LeMoyne, É., Léger, P.-M.: Expertise as a Mediating Factor in Conceptual Modeling. In: Information Systems and Neuroscience. pp. 85-92. Springer, (2018) Gegenfurtner, A., Lehtinen, E., Säljö, R.: Expertise Differences in the Comprehension of Visualizations: a Meta-Analysis of Eye-Tracking Research in Professional Domains. Educational Psychology Review 23(4), 523-552 (2011)

Gemino, A., Wand, Y.: A framework for empirical evaluation of conceptual modeling techniques. Requirements Engineering 9(4), 248-260 (2004)

Gemino, A., Wand, Y.: Comparing mandatory and optional properties in conceptual data modeling. In: Proceedings of the Tenth Annual Workshop on Information Technologies and Systems, Brisbane 2000, pp. 97-102 (2000)

Gemino, A., Wand, Y.: Evaluating modeling techniques based on models of learning. Communications of the ACM 46(10), 79-84 (2003)

Glaser, R.: Education and thinking: The role of knowledge. American psychologist **39**(2), 93 (1984)

Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs. International Journal of Industrial Ergonomics 24(6), 631-645 (1999)

Henderson, J.M., Hollingworth, A.: The role of fixation position in detecting scene changes across saccades. Psychological Science 10(5), 438-443 (1999).

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye tracking: A comprehensive guide to methods and measures. OUP Oxford, (2011)

Just, M.A., Carpenter, P.A.: Eye fixations and cognitive processes. Cognitive psychology 8(4), 441-480 (1976)

Krupinski, E.A.: The importance of perception research in medical imaging. Radiation medicine 18(6), 329-334 (2000)

Lurigio, A.J., Carroll, J.S.: Probation officers' schemata of offenders: Content, development, and impact on treatment decisions. Journal of Personality and Social Psychology 48(5), 1112 (1985)

Moody, D.: The “physics” of notations: toward a scientific basis for constructing visual notations in software engineering. IEEE Transactions on Software Engineering 35(6), 756-779 (2009)

Nickels, K.R., Curley, S.P., Benson, P.G.: Judgment-based and reasoning-based stopping rules in decision making under uncertainty, vol. 7285. University of Minnesota, (1995)

Parsons, J., Cole, L.: What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques. Data & Knowledge Engineering 55(3), 327-342 (2005)

Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological bulletin 124(3), 372 (1998)

Recker, J.C., Dreiling, A.: Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education. In: Toleman, Mark, Cater-Steel, Aileen, & Roberts, Dave (Eds.) 18th Australasian Conference on Information Systems, Toowoomba, Australia, pp. 356-366 (2007)

Reingold, E.M., Sheridan, H.: Eye movements and visual expertise in chess and medicine. Oxford handbook on eye movements, 528-550 (2011)

Shanks, G.: Conceptual data modelling: an empirical study of expert and novice data modellers. Australasian Journal of Information Systems 4(2) (1997)

Sheridan, H., Reingold, E.M.: Expert vs. novice differences in the detection of relevant information during a chess game: evidence from eye movements. Front Psychol **5**, 941 (2014)

Van Waes, L., Leijten, M., Quinlan, T.: Reading during sentence composing and error correction: A multilevel analysis of the influences of task complexity. Reading and Writing **23**(7), 803-834 (2009)

Yusuf, S., Kagdi, H., Maletic, J.I.: Assessing the comprehension of UML class diagrams via eye tracking. In: Program Comprehension, 2007. ICPC'07. 15th IEEE International Conference, pp. 113-122. IEEE (2007)

# Chapitre 3: Deuxième Article

## Through the Eyes of Expertise: Comparison of the Visual Attention of Experienced Business Analysts and Novices in Conceptual Modeling

Karl-David Boutin[1], Pierre-Majorique Léger[1], Christopher J. Davis[2], Alan R. Hevner[3], and Élise Labonté-LeMoyne[1]

[1] HEC Montréal, Québec, Canada
{karl-david.boutin, pierre-majorique.leger, elise.labonte-lemoyne}@hec.ca
[2]University of South Florida, Saint Petersburg, FL, USA
davisc@mail.usf.edu
[3]University of South Florida, Tampa, FL, USA
ahevner@usf.edu

### Abstract:

Business process models, or conceptual models, are widely used medium in design activities to facilitate communication about business domains and processes. Despite being an extensively researched topic, some aspects of conceptual modeling are yet to be fully explored and understood by academicians and practitioners alike. We study the attentional characteristics specific to experts and novices in a semantic and syntactic error detection task across 75 BPMN models. Our results suggest that experts tend to more falsely diagnose anomalies in error-free models than novices, while performing experts correctly distinguish more error-free stimuli than other participants. Furthermore, syntactic errors are diagnosed faster than semantic errors and error-free models are diagnosed the slowest. Our findings regarding the slight difference between experts and novices highlight the non-polarity of the concept of expertise and the need to further study how business analysts analyze conceptual models.

**Key words and phrases:** Conceptual modeling, process modeling, eye tracking, attentional characteristics, expertise, business process management, BPMN, business analysts, syntactic errors, semantic errors.

## Introduction

Business process modeling activities are used as a way to communicate and share knowledge, design and improve processes, and re-design decisions in organizations (Recker et al., 2012; Becker et al., 2004; Indulska et al., 2009). While always considered a key tool in information systems (IS), business process models are becoming more widely studied in the last decade (Recker et al., 2012; Indulska et al., 2009). However, evidence shows that despite the effort of the academic community, practitioners (e.g. business analysts) still struggle with some aspects of conceptual modeling, such as the standardization of modeling notations and methodologies (Indulska et al., 2009; Eikebrokk et al., 2008). Furthermore, the focus of the academic community does not always correspond with the practitioners' needs (Indulska et al., 2009). Areas of interest such as individuals' performance and practitioners' training seem to be of little interest to the research community compared to other elements, such as modeling grammar and method (Indulska et al., 2009; Jabbari et al., 2017; Wand and Weber, 2002). This creates a significant gap in the literature as the lack of knowledge on the nature of expertise in conceptual modeling is a hindrance to the improvement of business analysts' training curriculum (Davis et al., 2018).

The use of a visual notation, such as BPMN or UML, which has its own set of rules and constraints, allows the model to be understood by people across different departments and organizations (Davis et al., 2018). Yet, when using a visual notation, it can still be hard to fully grasp the model and it is almost impossible to design a flawless model without any ambiguities that will be understood by everybody the way it was meant to (Figl, 2017). In order to refine the notations and to improve the training of future business analysts, we need to understand how, on a cognitive level, business analysts read and comprehend graphical process models.

In order to better understand the expert modelers' heuristics, the main objective of the study is to probe the difference in cognitive processing between experienced business analysts and novices while reading and diagnosing errors in conceptual models. By exploring and understanding the differences between neophytes and more experienced modelers, we hope to highlight what can be considered 'best practices' in deciphering models and at the same time identify the limitations of visual notations. To do so, we use the concept of expertise to compare two groups of modelers in order to identify the skill-based adaptations that differentiate novice and expert designers (Davis et al., 2018). Understanding the heuristics of the experts in an error detection task will allow us to adapt the training curricula to facilitate the development of future business analysts.

Several researchers have explored the variations between novice and expert dichotomy in conceptual modeling (Shank, 1997; Yusuf et al., 2007; Recker and Dreiling, 2011; Koschmider et al., 2015). However, we are aware of no studies exploring the repertoire of skills, or 'competencies', outlined in Basselier et al. (2003), to assess their interdependence and capacity to differentiate novices from more experienced business analysts. The skills and abilities that differentiate experienced business analysts from novices, covering a broad spectrum, make the identification of success factors in conceptual modeling rather difficult.

This research strives to deepen understanding of the sometimes simplistic expert-novice dichotomy evident in prior studies, using eye-tracking instruments to capture an objective measurement of the actual behavior of our business analysts via their visual attention.

Our efforts to more fully articulate the expert-novice dichotomy also strives to identify and ameliorate limitations in the literature. Specifically, in this work, we considered semantic and syntactic error detection tasks by scrutinizing the differences between successful anomaly detection and unsuccessful diagnostics.

Our results suggest few statistically significant differences between novices and experts. Experts tend to detect more non-existent anomalies (false positives) in error-free models than novices. However, performing experts correctly diagnose error-free stimuli more efficiently. Syntactic errors tended to be diagnosed more quickly than semantic errors and models without any error generally took more time to diagnose than other models.

The paper is organized as follows: first, a literature review of the main concepts, starting by exploring the concept of conceptual modeling, before examining prior studies regarding expertise, and finally delving into the use of eye-tracking devices to capture visual attention. Then, the methodology, instruments, and measures used in the experiment are explained. The results are then listed and analyzed, before concluding the paper with insights for research and practice.

# Literature Review

## Conceptual Modeling

Conceptual modeling is a complex activity (Wand and Weber, 2002), essential to the design of IT artifacts (Davis et al., 2018). More than just a tool to facilitate comprehension of business processes (Figl, 2017), conceptual models are used, among other things, as a communication medium between users and developers, and to help business analysts understand business domains (Kung and Solvberg, 1986; Bavota et al., 2011; Moody, 2009; Parsons and Cole, 2005). They also play an important part in business process transformation, since they greatly facilitate the investigation of problems and limitations in organizations (Liberatore et al., 1998). Conceptual models are also used as a bridge between the business and IT focused actors, allowing them to understand each other easily and, thus, allowing them to work together on improving the business processes of the organization (Birkmeier et al., 2010).

By tapping into two-dimensional space, diagrams, or models, allow the organization of information by location, rather than having to follow a linear path like a textual representation (Larkin and Simon, 1987). This means that the relevant information is usually located in one place, which makes implicit information more obvious and models more concise than textual cases (Moody, 2009). This type of representation allows the business analysts to understand a situation, or problem, by crossing the diagram quickly, focusing on the different groups of information, rather than deciphering a text in their search for relevant elements (Larkin and Simon, 1987). Furthermore, the use of pictures has been shown to facilitate the acquisition and retention of information more readily than through the use of printed items.

Visual notations are composed of visual syntax, encompassing the visual vocabulary, which is the set of symbols, and the visual grammar, and visual semantics, which give meaning to the different symbols and to their relationship (Moody, 2009; Davis et al., 2018). However, while most studies concentrate on the effect of semantics on the comprehension of a model, for example by studying the level of abstraction of labels (Mendling et al., 2010; Mendling and Strembeck, 2008; Figl and Strembeck, 2015), few researchers have examined the effect of syntactic rules or offered syntactic guidelines

(Moody, 2009; Figl, 2017). This represents a significant gap in the literature, since syntactic differences between notations are as important, if not more prominent, than semantic variances (Moody, 2009).

The increasing popularity of process modelling in IS has spawned a significant number of notations and techniques to create conceptual models. This had the effect of increasing the number and disparity of academic and professional formations, each having to choose which notations to teach and how to actually teach it. Organizations also have to choose which notation they want for their process modeling and software supplier need the follow the demand and supply tools for the most popular notations (Recker and Dreiling, 2007). All those questions create a fertile environment for research. Furthermore, the lack of study comparing the differences between the semantic and syntactic components of visual notations, or simply the effect of the syntactic rules, present another opportunity to contribute to the literature.

While some experiments compared different notations or presentation mediums in order to identify the one having the most significant effect on comprehension (Yusuf et al., 2007; Ottensooser et al., 2012; Rodrigues et al., 2015; Recker and Dreiling, 2007), others have studied the effect of prior domain or modeling knowledge between users (Recker et al., 2014; Bera, 2012; Recker, 2013: Kummer et al., 2016). However, while there are recommendations on how to create better models, or how to adapt the models in function of your user's experience, few recommendations were made on how to improve training curriculums.

**The Evolving Nature of Expertise in Conceptual Modeling**

While the criteria to be considered an expert varies widely between fields and professions, since there is no consensus on the definition of expertise (Davis and Hufnagel, 2007; Wineburg, 1998), researchers tend to agree that, usually, experts are faster, more precise and more efficient than novices in their respective field (Sonnentag, 2000; Speelman, 1998). The main difference between novices and experts seems to be their organization of knowledge (Herbig and Büssing, 2004), whereas experts have more detailed and tightly connected schemata (Glaser, 1984; Lurigio and Carroll, 1985), which is defined by Glaser (1984) as the representation of the "knowledge that we experience—interrelationships between objects, situations, events, and sequences of events that normally occur" (Glaser, 1984, p. 100; Wineburg, 1998). Experts would then be able to infer other knowledge from the literal cues in a situation or a problem statement, where novices have less sophisticated strategies for using their knowledge to 'pick up' such subtle cues (Glaser, 1984; Lurigio and Carroll, 1985). Furthermore, the acquisition of a skill can bring changes to the brain, both by modifying the area of activation when processing a stimulus, to morphological changes increasing the grey matter dedicated to processing the type of stimuli trained for (Hill and Schneider, 2006). Chess players and radiologists, among other professionals requiring improved perceptual-motor skills, will have a higher performance using lower processing levels than novices, allowing them to perform more difficult discrimination tasks. Per contra, novices tend to use high-level processing, based more on generalizations (Hill and Schneider, 2006).

Evaluating expertise in the conceptual modeling environment is more complex than it seems since working with models require two different kinds of expertise: domain expertise, or expertise related to the semantic component of the models, and modeling expertise, or 'syntactic' expertise. An expert modeler, well versed in the creation of models using visual notation, may find quite hard to understand a model depicting a process from a domain that he doesn't know of, that he doesn't have any prior knowledge on. The opposite is also true; an expert in a domain may have some difficulties reading a conceptual model if he doesn't know the meaning of the symbols or if he is not used to working with models, even if the process depicted is well known to him.

In prior studies, a multitude of variables has been used to define modeling expertise, or expertise regarding the syntactic component of models, between groups. For example, self-reported measures on modeling familiarity (Weitlaner et al., 2013; Reijers and Mendling, 2011), frequency of work with models (Mendling and Strembeck, 2008) or objective measures of modeling knowledge (Recker, 2013; Figl et al., 2013; Mendling and Strembeck, 2008) have been used to compare groups of more experienced modelers, or 'experts', with novices. Across all of those studies and measures used, the frequency of use of flowcharts, prior experience with conceptual modeling (e.g. number of models created or read) and prior training had significant effect on model comprehension, where self-reported measures of knowledge and prior familiarity with modeling didn't differentiate the participants' comprehension (Figl, 2017).

In conceptual modeling, domain or 'semantic' expertise is usually assessed using self-reported measures on perceived domain knowledge (Figl, 2017). Across the experiments that studied the effect of prior domain knowledge on comprehension or performance, no significant effect has been observed.

Rather than studying the difference between expert modelers and novices, where expert modelers have been described in prior experiments with having at least four years of experience as modelers and had contributed to the development of at least ten conceptual models (Shanks, 1997), we focus our attention on the business analysts. Indeed, nowadays the majority of business analysts have to work continuously with conceptual models, whether by creating or reading them, and thus, form the core of the practitioners.

Furthermore, in accordance with the concept of IT competence, as defined by Basselier et al. (2003), business analysts have more IT knowledge - which is the relevant knowledge and the capability to access more IT-related knowledge - and IT experience than novices (Basselier et al., 2003). Therefore, business analysts, by having more experience with IT projects and by possessing deeper understanding and IT-related knowledge, are better suited as practitioners to use and interact with conceptual models than novices (i.e. individuals who have limited or no experience in situations characteristic of their domain) without any significant IT-related experience or knowledge.

Moreover, Patel and Groen's distinction of 'specific' and 'generic' expertise, where 'generic' experts have generic knowledge of the domain and 'specific' expertise is linked with specialized knowledge of the domain, and definitions of the levels of expertise allow us to place the business analysts in the 'subexpert' group and the novices in the 'layperson' group (Patel and Groen, 1991; Wineburg, 1998). Indeed, the average business analyst having a generic knowledge of IS and conceptual modeling, by their background and formation, are not as specialized as experienced modelers, but still have more expertise than novices, which are only equipped with commonsense and everyday knowledge.

Thus, any business analysts are used as surrogate expert modelers, by proposing for the purposes of this research that their IT competences and 'generic' expertise differentiate them from novices, and thus refer to them as 'experts' in the remainder of the article.

**Visual Attention**

The use of eye-tracking to monitor the visual attention of participants has been tested and proven as an effective way to assess the moment-to-moment cognitive processing of visual stimuli (Bednarik and Tukiainen, 2006; Rayner, 1998; Yusuf et al., 2007). Evidence suggests that attention and saccades, which are the quick movements of the eyes between different locations (Yusuf et al., 2007), are closely linked (Rayner, 1998), while fixations are linked with the cognitive processing of visual information (Yusuf et al., 2007; Just and Carpenter, 1976; Zhan et al., 2016). Technological innovations made eye-tracking instruments more accurate and reliable, while removing the need to use intrusive goggles or headset to capture precise visual data (Lupu and Ungureanu, 2013).

Multiple eye-tracking studies have evaluated the eye movement of participants during an anomaly detection task. While most of those experiments used anomalous textual sentences (Braze et al., 2002; Ni et al., 1998; Zhan et al., 2016) or anomalies in radiography (Krupinski, 2000; Reingold and Sheridan, 2011) as visual stimuli to assess the variation in eye movements, the eye-tracking methodology is quickly gaining popularity in other domains, from forensics to art (Reingold and Sheridan, 2011).

These studies concluded that more fixations will land on the relevant information, which in our case is the anomalies, and that those fixations tend to become longer than the fixations on irrelevant information (Van Waes et al., 2009; Henderson and Hollingworth, 1999; Holmqvist et al., 2011).

Furthermore, the number of fixations is related to the effectiveness of the search (Holmqvist et al., 2011; Goldberg and Kotval, 1999), where a higher number of fixations usually report an ineffective search. Finally, the total view time of the stimuli has been found to be inversely related to the detection of anomalies (Van Waes et al., 2009). Indeed, a higher time spent on a stimulus is correlated with a lower chance of identifying the anomaly, possibly explained by more cognitive resource being drawn away for the encoding of the stimulus.

These findings offer a better understanding of the visual characteristics related to successful anomaly detection in the context of conceptual modeling, which allow us to propose our first three hypotheses:

*H1 — Successful error detections in conceptual modeling will require less time spent looking at the stimulus than unsuccessful error detections.*

*H2 — Successful error detections in conceptual modeling will require, in total, fewer fixations than unsuccessful error detections.*

*H3 — Successful error detections in conceptual modeling will require, on average, shorter fixation duration than unsuccessful error detections.*

Multiple experiments studied the difference in visual attention between experts and novices in different search tasks. Those studies allow us to have a better idea of how expertise influences visual characteristics in different domains, even though it is not recommended to generalize eye movements meaning across tasks or domains, since contextual demands and task complexity might greatly differ (Rayner, 1998; Gegenfurtner et al., 2011).

Overall, the main findings of prior work suggest that experts tend to spend less time on a stimulus, less and shorter fixations on average and have all around better performance than novices (Krupinski, 2000; Gegenfurtner et al., 2011; Reingold and Sheridan, 2011; Sheridan and Reingold, 2014). These results align with our hypotheses on the visual characteristics of successful error detection, which may be explained by the tendency of experts towards more efficient search strategies and consequently better performance (Recker and Dreiling, 2007; Yusuf et al., 2007). Accordingly:

*H4 — Experts in conceptual modeling will spend less time looking at the stimulus than novices.*

*H5 — Experts in conceptual modeling will require, in total, fewer fixations than novices.*

*H6 — Experts in conceptual modeling will require, on average, shorter fixation duration than novices.*

*H7 — Experts in conceptual modeling will diagnose the anomalies more accurately than novices.*

## Method

### Participants

A within-subject experiment was conducted in order to test our hypotheses. 30 participants (15 males, 15 females, Age avg. = 28.63) were recruited and manually divided into two groups. The sample was screened to only allow participants who weren't diagnosed with any neuropsychological conditions or major vision problems that will require glasses to use a computer, in order to meet our instruments constraints. The research was approved by our institution's Research Ethics Board (REB), and each participant signed a consent form and received a small monetary compensation from the university bookstore.

The 'Novice' group was composed of 15 participants (9 males, 6 females) and were recruited among the volunteers enrolled in our institution's panel. This group's participants were between 21 and 38 years old (Avg. = 24; SD = 4.06612). Any participant with a background in IT and business analysis was excluded, leaving only those that had never used or learned any visual notation. Even though this group was mostly composed of undergraduate and graduate students, which might weaken perception of the external validity of the study, the use of students over practitioners allowed us to control the prior technique and domain knowledge of the participants in order to make sure that our novices have undeveloped IT competencies (Recker and Dreiling, 2007; Gemino and Wand, 2004; Batra et al., 1990).

The remaining fifteen (15) participants (6 males, 9 females) comprised our 'Expert' group. They were recruited, in part, at an International Institute of Business Analysis (IIBA) convention that took place in a major North American city in February 2018. The ages of this group range from 22 to 53 years (Avg. = 33.26; SD = 9.862161). The experts had to be business analysts and to have worked on at least 1 project using conceptual modeling, for a minimum of 15 hours of work. Since each organization can develop their own 'flavor' of BPMN or other notation, we also made sure to recruit experts from different institutions, in order to minimize the risk that participants will be biased by practices specific to their organization.

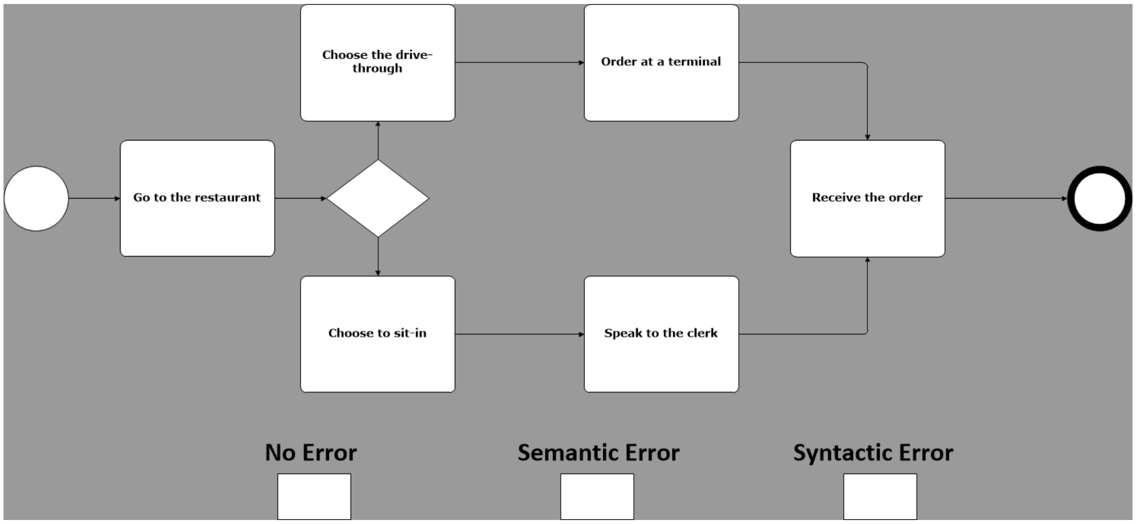
**Experimental Stimuli**

Since conceptual modeling covers such a large scale of notations and domains, we limited our choice of formalism to notations used in process modeling, such as UML and BPMN.

The Business Process Modeling Notation, or BPMN, is an international standard for business process notation published by the Business Process Management Initiative (now Object Management Group) in 2004. BPMN strives to be understandable by all business users, from business analysts creating the models to business actors using or monitoring the processes and even developers (Birkmeier et al., 2010). Among the analytical evaluation studies, Wahl and Sindre (2005) found that BPMN is easy to learn for simple use, even though it can be more complex than other notations when used with advanced modeling concepts (Wahl ad Sindre, 2005). However, empirical experiments have found no evidence that the use of BPMN over another notation would significantly improve the comprehension of the participant (Sandkuhl and Wiebring, 2015; Jost et al., 2016; Birkmeier et al., 2010). Furthermore, the growing popularity of BPMN in the commercial and academic settings lead to its selection as the visual notation used in this experiment.

Building upon the experimental design proposed by Davis et al. (2018), a laboratory experiment was conducted using a 5x5x3 within-subject design, where five (5) complete scenarios, or 'blocks' were created, each of them containing 5 incomplete and smaller versions of the scenario, or 'sentences' (Davis et al., 2018). Each sentence was then further manipulated to presents 3 models: one with no known errors (Figure 1.); one with a known semantic error (Figure 2.), and one with a known syntactic error (Figure 3.) (Davis et al., 2018). A total of 75 models, 15 models per scenario and 25 for each error type, were created. Boxes indicating the three error types were added at the bottom of each model and were used by participants to indicate their diagnosis.

Syntactic errors include the use of invalid symbols (Davis et al., 2018) (e.g. the use of a BPMN 'start event' symbol to represent a 'gateway') or a non-consistent flow (e.g. a misdirected flow between two activities). Semantic errors, on the other hand, are subtler and cannot be identified at a glance, in addition to being difficult to recognize by a compiler or other verification technologies, making them especially costly and hard to correct (Davis et al., 2018; Dijkman et al., 2008). By using valid symbols but ambiguous




design (e.g. sequence of activities in a scenario mis-ordered), they present an unintended and puzzling message.

**Fig. 1. Example of a model with no error**

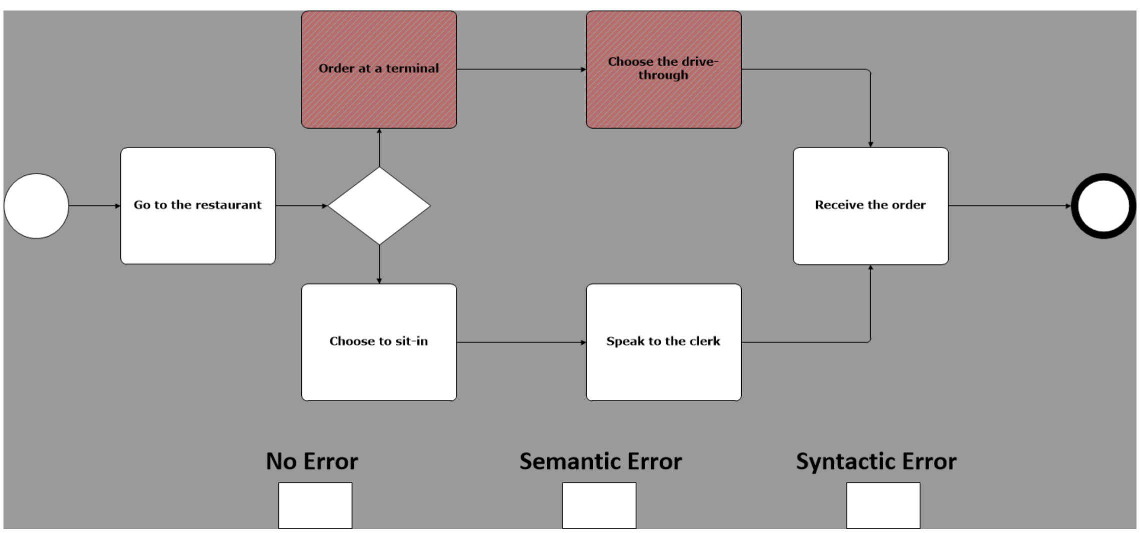



**Fig. 2. Example of a model with a semantic error (mis-ordered activities)**

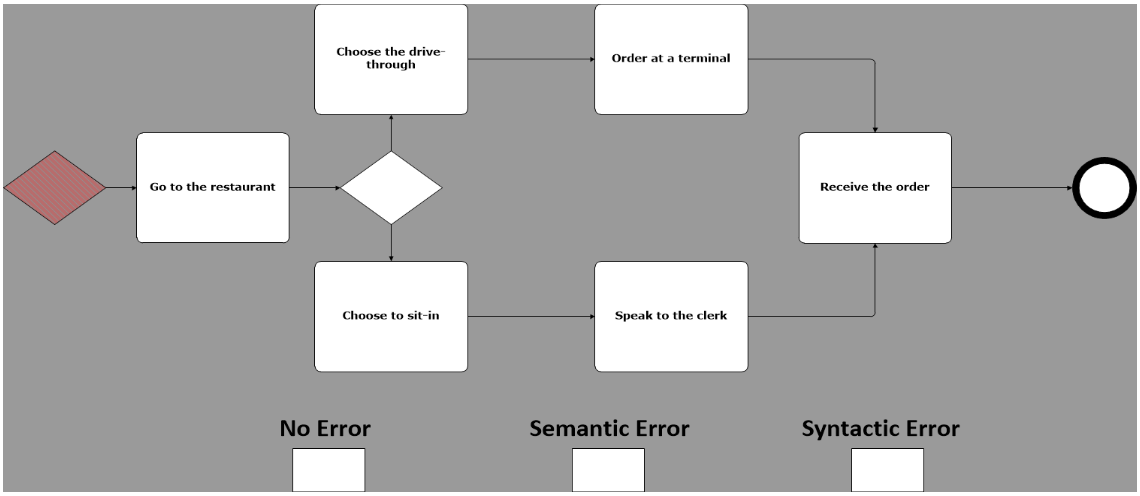



**Fig. 3. Example of a model with a syntactic error (use of invalid symbols)**

While several studies found that the domain knowledge had no significant effect on the understanding of the models in a model comprehension task (Recker and Dreiling, 2007; Bera, 2012; Recker et al., 2014; Turetken et al., 2016), we feared that the disparity in the prior knowledge of the business domains of the models between the participants would greatly influence the ease with which they would pinpoint errors in an anomaly detection task (Gemino and Wand, 2004; Birkmeier et al., 2010). Therefore, we based our models on simple and well-known scenarios to all participants (e.g. fast food ordering process). By limiting the range and quantity of symbols used between 8 and 13 elements per model, we controlled the complexity of each scenario and models (Sanchez-Gonzalez et al., 2010). In accordance with prior BPMN literature (Wahl and Sindre, 2005), only the basic symbols were used, since the use of more advanced components of the notation greatly complicates the comprehensibility of the models. Exclusive gateways were used, since the use of gateways has a positive effect on the comprehension of a model (Recker, 2013), but we did not use any of the other types of gateway since a use of a heterogeneous range of gateways tends to lower the comprehension of the models (Sanchez-Gonzalez, 2010). By combining the use of readily comprehensible models with a training presentation and some practice tasks, we partly mitigated the effect of the variation of prior technical knowledge between each participant (Bavota et al., 2011).

The training consisted of a PowerPoint presentation explaining the symbols used and the two types of errors present in the experiment (i.e. semantic and syntactic errors). The participants went through the presentation at the beginning of the experiment, after the calibration of the instruments. They were allowed to take as much time as needed. The content of the presentation was tested on 6 participants with no, or close to no, experience with BPMN. After reading the presentation, the pretested participants were asked to describe the different symbols and rules explained in the training. The stimuli were then improved, and any confusion removed.

After the training presentation, the participants were given a practice exercise. Just like the experiment, the practice consisted of identifying and diagnosing an error in a conceptual model. The task was composed of three (3) models, where each one had a different type of error (i.e. no error, semantic error, and syntactic error). To avoid any form of bias, the practice models were not related to the scenarios used later in the experiment and the modeling (experimental) environment was the same (Bavota et al., 2011). The only difference between the training task and the experiment was that the participant could see the correct answer after each practice model. This way, any remaining confusion regarding the error types was ameliorated before the experiment.

The training presentation and the practice task allowed us to make sure every participant had the necessary knowledge to complete the experiment and to mitigate the effect of learning through trial (Bavota et al., 2011). While useful with novices and experts who were less familiar with BPMN, the training phase of the experiment also allowed us to make sure that the experts accustomed with the use of BPMN were still using the normalized rules of BPMN and were not biased by some of their own organization's standards.

To complete the tasks, the participant had to manually click on the modeling error and on the box classifying the error type at the bottom of the screen, using the mouse (see Figure 4). After each click on the model, a visual indicator would be placed on the location of the click and would disappear after 0.5 seconds. This indicator was used to provide visual feedback to the participant that the click was registered, in order to mitigate any confusion about the user's actions. If the model contained no error, the participant simply had to click on the box indicating "No Error". After identifying the error, by clicking on it, and the error type, by clicking on the corresponding box, participants manually advanced to the next model by pressing the spacebar. After completion of a scenario, the researcher opened the online questionnaire and instructed the participant to complete it. The researcher then closed the questionnaire and started the next set of models, or ‘scenario’.

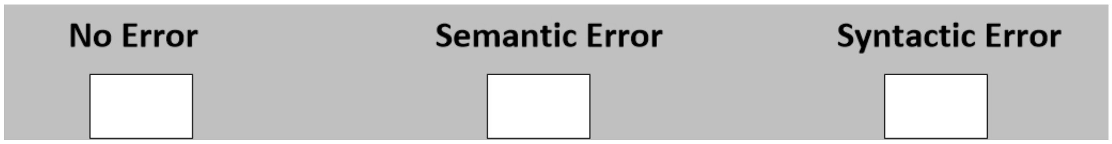



**Fig. 4. Boxes indicating error types**

**Protocol**

As the participants arrived at the laboratory lobby, the researcher greeted them and explained roughly the stages of the experiment. They were then asked to read and sign the consent form, while the experimenter made sure that the equipment was ready to run and that all the required software was initiated. After the consent form was signed by both the participant and the experimenter, participants were taken to the laboratory and the eye-tracking device calibrated. These steps took approximately 10 minutes.

The participant then went through the training presentation and practice task, which took, on average, between 5 and 7 minutes to complete. On completion of the practice task, the participants started their first task for a random scenario. For each scenario, the participants had to identify and diagnose errors in 15 models, shown in a random order, without any time limitations. After completing a scenario, the participants would then start another error detection task, for another scenario at random. The experiment would conclude when a participant would go through the 5 scenarios, totaling the 75 models, and a questionnaire on their previous experience with conceptual modeling.

At the end of the experiment, which took around 45 minutes, the participant, was given their compensation and escorted back to the building lobby.

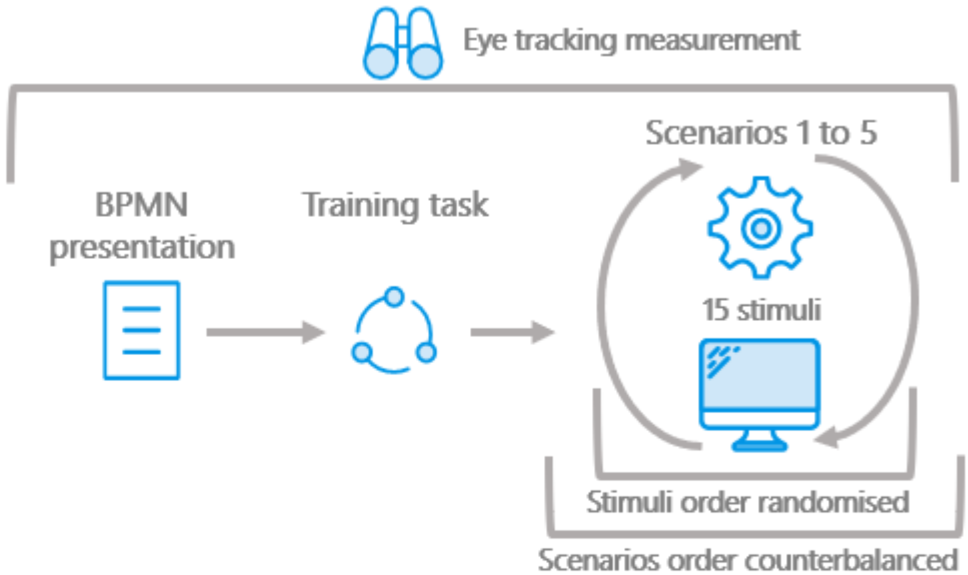



**Fig. 5. Experimental Design**

**Measures**

We captured the behavioral measures, which translate into eye movements, using SMI RED 250 eye-tracker (Red 250, SensoMotoric Instruments GmbH, Teltow, Germany). The instrument was configured at a sampling frequency of 60 Hz and a fixation duration threshold of 200 ms (Holmqvist et al., 2011; Rayner, 1998). Following the calibration, using a 9-point predefined calibration grid, the eye-tracker was adjusted for each participant, to a gaze-position deviation of 0.5◦ or less.

For each model, areas of interest (AOIs) were mapped to the location of the error. Additional margins of at least 1.5◦ were added to the AOIs to mitigate the imprecision of the eye-tracker (Holmqvist et al., 2011). Figure 6 shows an example of a model with AOIs, where the AOI can be seen on the syntactic error. These AOI mappings allowed us to gather data on the proportion of fixation and time to first fixation on precise locations in our models.

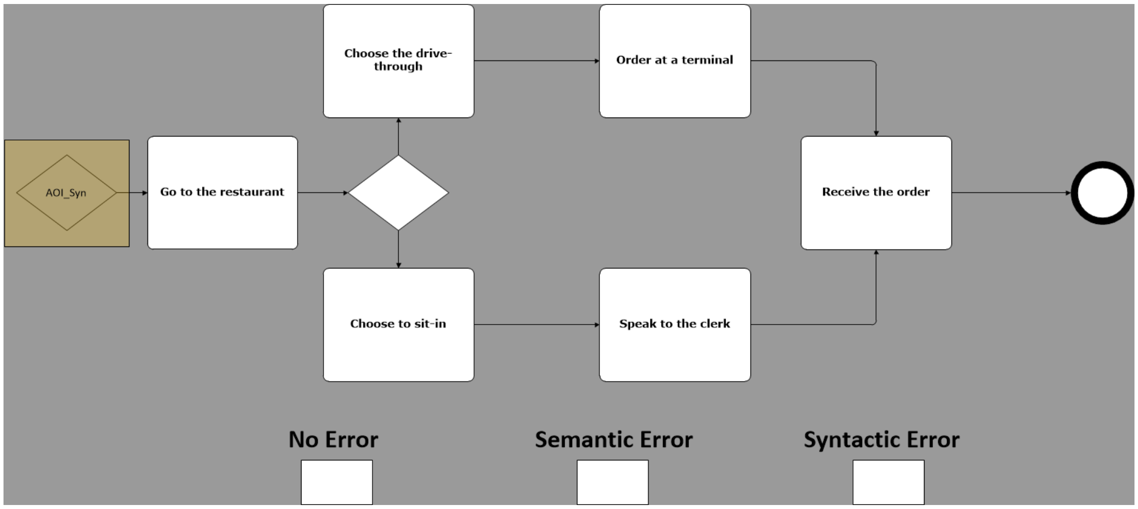


**Fig. 6. Example of a model with visible AOI**

As the literature tends to agree that the cognitive processing of visual stimuli is done during fixations (i.e. the stabilization of the eye on an object) (Yusuf et al., 2007; Just and Carpenter, 1976; Zhan et al., 2016), we gathered the fixations, and their duration, inside each area of interest (AOI). We also took into account the total view time of each stimulus, in order to evaluate the response time of the participants.

The experiment was set up using SMI Experiment Center 3.7.56 and the data were processed and analyzed using SMI BeGaze 3.7.40 software. The stimuli were created using Microsoft Visio 2010. The statistical analysis was carried out using Stata/MP 15.1.

The error rate for each participant was also determined. A performance score was created for the three types of error and calculated by manually reviewing, with the help of the experiment's recordings, each answer given by the participants. Those scores, in percentage, were used to identify which type of error was the hardest to diagnose and for which error group the difference between experts and novices was the largest.

A questionnaire was created in order to determine the prior experience of the participants. It was composed of 7 questions assessing the number and kinds of visual notations known by the participant, the number of projects and hours spent working on conceptual models and the kind of manipulation done in those projects. In accordance with literature on expertise, rather than using the amount of time spent as a business analyst as our indicator of modeling expertise, since experience in itself is often a poor predictor of true expertise (Gobet, 2016), we chose objective measures of modeling experience, being the number of projects involving conceptual models in which the participant took part and the hours spent working with conceptual models, in BPMN or any other notations, to define our experts and novices.

## Analysis

To test our hypotheses, linear regressions with mixed model and a two-tailed level of significance were performed. Dummy variables were created to represent the error types: dErrSem for semantic errors, dErrSyn for syntactic errors with invalid symbols (ErrSyn2) and with non-consistent flow (ErrSyn3), dErrSyn2 only for syntactic errors with invalid symbols and dNoError for stimuli with no error. Syntactic errors with non-consistent flow (ErrSyn3) are isolated when dErrSyn = 1 and dErrSyn2 = 0 and, therefore, no dummy variable dErrSyn3 variable was created. The binary variable Expertise was also created to distinguish our two groups of participants, where Expertise = 1 when the participant is a business analyst. The variable dWhiteSpace regroups everything that is not inside an AOI and is considered as 'irrelevant' information. For each stimulus with an error (thus excluding stimuli with dNoError), there will be a dummy variable (dErrSem, dErrSyn or dErrSyn2) representing the error area and dWhiteSpace representing the rest of the stimulus.

By creating a median split on the overall performance of experts, we can create and compare two groups: the performing and underperforming experts. This manipulation allows us to push our analysis further, and to articulate the heuristics of experts with good and poor performance. The dummy variable 'dGoodExpert' was then created, where dGoodExpert = 1 represents the group of high performing experts. A similar manipulation was carried out by creating a median split on the performance of all participants, thus creating dPerformance, where dPerformance = 1 represents the group of high performing participants, novices and experts alike. The results of the linear regressions and correlation are shown in the tables in the following section.

## Results

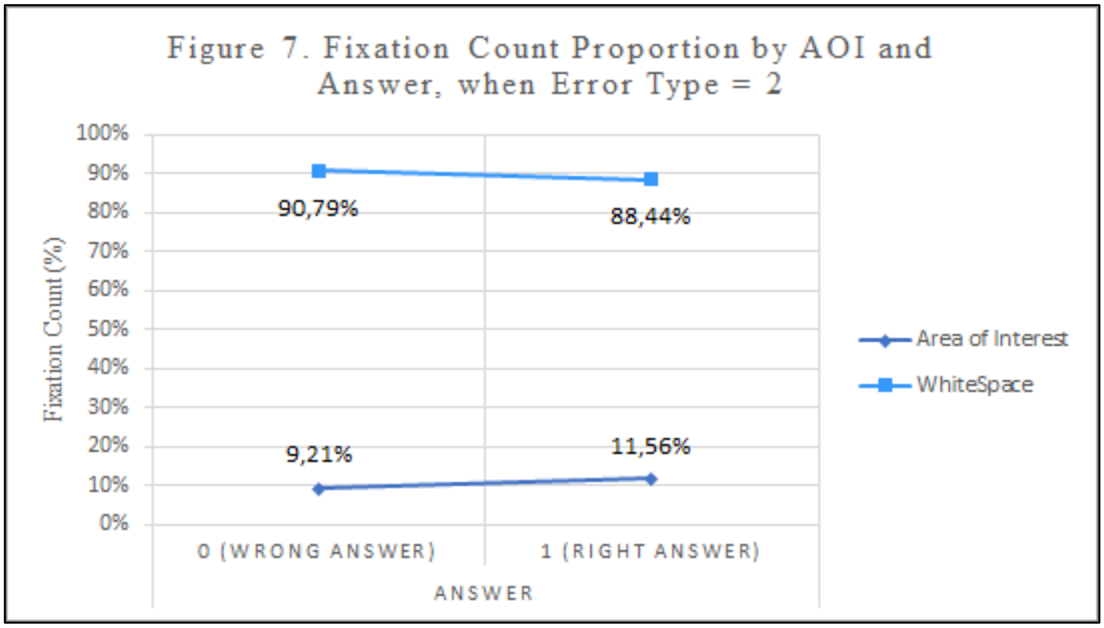
The first hypothesis (H1) states that successful identification and diagnosis of errors in conceptual models will take less time than unsuccessful answers. We compared the effect of the variable ‘Answer’ (i.e. if the participant successfully diagnosed the error, Answer = 1, if the diagnostic was wrong, Answer = 0) on the measure ‘Total View Time’. A significant relationship between those two variables was found ($b = -0.4143$, $p<0.001$), meaning that correct diagnostics tend to be looked at for a significantly shorter amount of time than wrong answers, thus supporting the hypothesis. Furthermore, Table 1 shows that syntactic errors tend to be diagnosed faster than other error types ($b = -0.3237$, $p<0.001$) and, more specifically, stimuli with syntactic errors that used invalid symbols (ErrSyn2) are diagnosed faster than stimuli with non-consistent flow (ErrSyn3) ($b = -0.3258$, $p<0.001$). These results contrast with those for the ‘No Error’ group, which were looked at longer than other error types ($b = 0.2913$, $p<0.001$). No statistically significant result was found for semantic errors.

While H2 states that successful error detection will require fewer fixations than incorrect diagnostics, a barely significant result suggests the contrary. Indeed, ‘Answer’ seems to have a hardly significant positive relationship, if considered significant at all, with ‘Fixation Count’ ($b = 0.1673$, $p<0.0864$), implying that good answers are linked with a higher fixation count. However, while the fixation count in the AOI is lower than in the White Space (i.e. everything that is not inside the AOI) for both correct and incorrect diagnostic, the proportion of fixation in the relevant area is higher for the accurate diagnostics of syntactic errors ($b = 0.0236$, $p<0.003$) as we can see in Figure 7.

**Table 1. Results of linear regressions**

| | Sex | LogAge | Expertise | dPerformance | dGoodExpert | dNoError | dErrSem | dErrSyn | dErrSyn2 | Answer |
|---|---|---|---|---|---|---|---|---|---|---|
| Perf_Total | -0.0674* | -0.0911 | 0.0052 | 0.1193*** | 0.0941* | 0.0001 | -0.0001 | 0.0001 | -0.0014 | 0.0602**** |
| | (0.0380) | (0.0807) | (0.0389) | (0.0352) | (0.0491) | (0.0001) | (0.0001) | (0.0001) | (0.0014) | (0.0128) |
| Total view time | 0.2373*** | 0.3843** | 0.0831 | -0.0648 | 0.0834 | 0.2913**** | 0.0325 | -0.3237**** | -0.3258**** | -0.4143**** |
| | (0.0724) | (0.1553) | (0.0821) | (0.0836) | (0.1351) | (0.0404) | (0.0316) | (0.0460) | (0.0501) | (0.0314) |
| Time to first fixation | 0.0599 | -0.1358 | -0.0247 | 0.1912 | 0.1855 | -3.7968**** | 1.8252**** | 2.0672**** | -0.1922* | -0.2053 |
| | (0.1157) | (0.2955) | (0.1170) | (0.0463) | (0.1782) | (0.1215) | (0.0785) | (0.0754) | (0.1082) | (0.1546) |
| Fixation count | 0.1664** | 0.2220 | 0.1168* | 0.0152 | 0.1005 | 2.2335**** | -0.7364**** | -1.4956**** | -0.2820**** | 0.1673* |
| | (0.0629) | (0.1380) | (0.0681) | (0.0463) | (0.1191) | (0.0473) | (0.0295) | (0.0405) | (0.0520) | (0.0942) |
| Fixation duration (ms) | 10.9187*** | 15.5805** | 6.2263 | -2.3073 | 2.5698 | 1.02e+02**** | -4.09e+01**** | -6.09e+01**** | -7.8289**** | 1.6253 |
| | (3.2504) | (6.7133) | (3.6720) | (3.7416) | (6.2372) | (3.6638) | (1.8398) | (2.3814) | (1.5880) | (4.4275) |
| Fixation duration (%) | -0.0423 | -0.1289 | -0.0040 | 0.0298 | 0.0159 | 1.9532**** | -0.7196**** | -1.2322**** | -0.0044 | 0.6760**** |
| | (0.0524) | (0.1151) | (0.0530) | (0.0539) | (0.0948) | (0.0410) | (0.0336) | (0.0612) | (0.0663) | (0.0922) |
| Perf_NoError | -0.1276*** | -0.1876 | -0.0942* | 0.1483*** | 0.1921** | - | - | - | - | - |
| | (0.0446) | (0.1134) | (0.0470) | (0.0476) | (0.0655) | | | | | |
| Perf_Sem | 0.0025 | 0.0455 | 0.0415 | 0.0962** | 0.0110 | - | - | - | - | - |
| | (0.0478) | (0.0835) | (0.0462) | (0.0470) | (0.0540) | | | | | |
| Perf_Syn | -0.0771 | -0.1309 | 0.0680 | 0.1134* | 0.0797 | - | - | - | - | - |
| | (0.0647) | (0.1407) | (0.0633) | (0.0626) | (0.0815) | | | | | |
| Perf_Syn2 | 0.0470 | -0.0909 | 0.0330 | 0.0453 | -0.0084 | - | - | - | - | - |
| | (0.0769) | (0.1795) | (0.0776) | (0.0792) | (0.1163) | | | | | |
| Perf_Syn3 | -0.1601** | -0.1571 | 0.0912 | 0.1605** | 0.1409 | - | - | - | - | - |
| | (0.0681) | (0.1546) | (0.0695) | (0.0680) | (0.0883) | | | | | |
| Answer | -0.3469* | -0.4304 | 0.0417 | 0.6172**** | 0.4944** | 0.2365 | -0.4336*** | 0.2191 | 0.0871 | - |
| | (0.1974) | (0.3991) | (0.2061) | (0.1804) | (0.2395) | (0.1658) | (0.1357) | (0.1685) | (0.2088) | |
| Input | 0.5618 | 0.5629 | 0.6165* | -0.5169 | -0.9732** | 4.3134**** | -1.4103**** | -0.4654** | -1.1883**** | - |
| | (0.3489) | (0.6923) | (0.3288) | (0.3470) | (0.4033) | (0.7370) | (0.2027) | (0.1923) | (0.3442) | |

Notes: Standard errors in parentheses; signif.: **** = p<0.001, *** = p<0.01, ** = p<0.05, *=p<0.1




Similar results were found for semantic errors ($b = 0.0062$, $p<0.022$) and syntactic errors of non-consistent flow ($b = 0.0196$, $p<0.008$).

No statistically significant results were observed for H3, which states that successful error diagnosis will be linked with shorter fixation duration.

**Table 2. Effect of errors type on attentional characteristics, when Expertise = 1**

| | dWhiteSpace | dErrSem | dErrSyn | dErrSyn2 |
|---|---|---|---|---|
| Total view time | 0.3716****<br>(0.0455) | 0.0457<br>(0.0423) | -0.4167****<br>(0.0652) | -0.3189***<br>(0.0921) |
| Fixation count | 1.8606****<br>(0.0629) | -1.0475****<br>(0.0499) | -1.7427****<br>(0.0681) | -0.3491****<br>(0.0606) |
| Fixation duration (ms) | 80.8083****<br>(4.3135) | -5.12e+01****<br>(3.1097) | -7.00e+01****<br>(3.9850) | -1.04e+01****<br>(1.6075) |
| Fixation duration (%) | 1.7238****<br>(0.0811) | -1.0859****<br>(0.0440) | -1.4991****<br>(0.1128) | -0.1308<br>(0.0770) |

Notes: Standard errors in parentheses; signif.: **** = p<0.001, *** = p<0.01, ** = p<0.05, *=p<0.1

H4, which states that experts in conceptual modeling will spend less time looking at the stimulus than novices, is not supported since no statistically significant results were found linking ‘Expertise’ with ‘Total View Time’. However, Table 2, which combines statistical analysis of a dataset when only including the experts (Expertise = 1), suggests that among experts, syntactic errors tend to be diagnosed faster than other error types ($b = -0.4167$, p<0.001), with the ‘invalid symbols’ group being diagnosed faster than the ‘non-consistent flow’ errors (b = -0.3189, p<0.01), and where stimuli without any errors were diagnosed slower (b = 0.3716, p<0.0038). These results concur with the findings for H1. Indeed, just like accurate diagnostics, experts tend to identify syntactic errors faster and diagnose error-free stimuli slower. Again, no statistically significant result was found for semantic errors.

Table 3 shows the significant results from statistical analysis from a dataset only including the high performing experts (dGoodExpert = 1). A similar pattern can be found, where, among these proficient experts, syntactic errors are diagnosed faster than other types of errors (b = -0.4884, p<0.0013), with faster identification of errors involving invalid symbols usage rather than non-consistent flow (-0.3489, p<0.008) and a longer total view time with stimuli in the ‘No Error’ group (b = 0.3683, p<0.0005). However, a statistically significant link was found for the semantic errors, where the high performing experts tend to respond more slowly to stimuli with semantic errors than other error types.

**Table 3. Effect of errors type on attentional characteristics, when dGoodExpert = 1**

| | dWhiteSpace | dErrSem | dErrSyn | dErrSyn2 |
|---|---|---|---|---|
| Total view time | 0.3683****<br>(0.0622) | 0.1201*<br>(0.0607) | -0.4884***<br>(0.0955) | -0.3489***<br>(0.0953) |
| Fixation count | 1.8245****<br>(0.0629) | -1.0295****<br>(0.0830) | -1.7074****<br>(0.0861) | -0.4257***<br>(0.0847) |
| Fixation duration (ms) | 80.3340****<br>(5.3879) | -5.01e+01****<br>(4.7160) | -7.04e+01****<br>(4.3249) | -1.23e+01***<br>(2.5090) |
| Fixation duration (%) | 1.6751****<br>(0.1096) | -1.1033****<br>(0.0773) | -1.4095****<br>(0.1364) | -0.1725<br>(0.0944) |

Notes: Standard errors in parentheses; signif.: **** = p<0.001, *** = p<0.01, ** = p<0.05, *=p<0.1

H5 and H6 were found to be inconclusive since no statistically significant result was observed. The effect of ‘Expertise’ on ‘Fixation Count’ (b = 0.1168, p<0.0970) and ‘Fixation Duration’ (b = 6.2263, p<0.1007), even if not significant, seems to contradict our hypotheses.

H7, which proposes that experts should diagnose anomalies more accurately than novices, was tested by analysing the effect of ‘Expertise’ on the ‘Total Performance’ and the individual performance for each error type (see Table 1). The only modestly significant link is found for the performance with the ‘No Error’ group of stimuli, where, contrary to our expectations and hypothesis, experts tend to have a lower performance for error-free stimuli than novices (b = -0.0942, p<0.0547).

To get a deeper understanding of the relationship between our classification of expertise and the responses to the stimuli, we analyzed the effect 'Expertise' on the variable 'Input' for wrong diagnostics, where Input = 1 means that the participant diagnosed a wrong anomaly and Input = 0 denotes that the participant wrongly thought that the stimuli didn't have any error. Results suggest that experts tend to diagnose wrong anomalies more than novices (b = 0.6165, p<0.0608). These false (or secondary) positives were unexpected and highlight the imperfectability of our experimental design. Like error free code, completely unambiguous BPMN models are a utopian myth. Although clearly a limitation, the paradox evident here prompt important questions about the soundness of the expert-novice dichotomy.

Contrary to our expectations, substituting 'Expertise' with 'dGoodExpert' showed that high performing experts tend to diagnose that stimuli didn't have any errors more frequently than underperforming experts (b = -0.9731, p<0.0158), suggesting that they diagnose more 'false negatives'. However, high performing experts are also associated with a higher success rate with error-free stimuli than other participants (b = 0.1921, p<0.0109).

The paradox is further highlighted when we isolate the error types of the stimuli wrongly diagnosed. We can see in Table 4 that experts, when offering a false diagnosis, tend to identify false anomalies in syntactic errors with invalid symbols use (b = 1.2481, p<0.0048) and syntactic errors with non-consistent flow (b = 0.9518, p<0.0844) more than novices. Despite this finding, the effect of 'Expertise' on 'Input' for semantic errors was positive, but statistically insignificant (b = 0.2152, p<0.5858), while high performing experts are more inclined to inappropriately respond 'No Error' than underperforming experts (b = -1.1518, p<0.0187).

**Table 4. Effect of expertise on the type of error answered, when Answer = 0**

| | Expertise | dGoodExpert |
|---|---|---|
| Input | 0.6165*<br>(0.3288) | -0.9733**<br>(0.0158) |
| Input<br>(Error type = 1) | 0.2152<br>(0.3948) | -1.1518**<br>(0.4895) |
| Input<br>(Error type = 2) | 1.2481***<br>(0.4425) | 0.6205<br>(0.5549) |
| Input<br>(Error type = 3) | 0.9518*<br>(0.5517) | -0.7816<br>(1.0133) |

Notes: Standard errors in parentheses; signif.: **** = p<0.001, *** = p<0.01, ** = p<0.05, *=p<0.1

Contrary to our expectations, our analysis shows that, prior experience is negatively correlated with performance. For instance, Nb_Q2, which is the number of notations previously experienced, has a negative effect on performance (b = -0.3410, p<0.0703). This infers - again, paradoxically, that participants with experience of more notations beforehand tend to have a <u>lower</u> score. Similarly, lQ3_cont and lQ6_cont, which represent the number of projects in which they used any visual notation (lQ3_cont) or specifically BPMN (lQ6_cont), both have a <u>negative</u> impact on performance (b = -0.3991, p<0.0320) and (b = -0.3294, p<0.0810) respectively.

**Table 5. Correlation table of items of Past Experience questionnaire**

| | dQ2 | Nb_Q2 | lQ3_cont | lQ4_cont | lQ6_cont | lQ7_cont |
|---|---|---|---|---|---|---|
| Performance | -0.2183 | -0.3410* | -0.3991** | -0.2828 | -0.3294* | -0.2245 |
| dGoodExpert | 0.2774 | -0.2206 | -0.3480 | -0.2122 | -0.0255 | 0.1168 |

Notes: Signif.: **** = p<0.001, *** = p<0.01, ** = p<0.05, *=p<0.1

**Table 6. Summary of hypotheses**

| Hypothesis | Description | Conclusion |
|---|---|---|
| **H1** | Successful error detections in conceptual modeling will require less time spent looking at the stimulus than unsuccessful error detections. | Supported |
| **H2** | Successful error detections in conceptual modeling will require, in total, fewer fixations than unsuccessful error detections. | Not supported |
| **H3** | Successful error detections in conceptual modeling will require, on average, shorter fixation duration than unsuccessful error detections. | Not supported |
| **H4** | Experts in conceptual modeling will spend less time looking at the stimulus than novices. | Not supported |
| **H5** | Experts in conceptual modeling will require, in total, fewer fixations than novices. | Not supported |
| **H6** | Experts in conceptual modeling will require, on average, shorter fixation duration than novices. | Not supported |
| **H7** | Experts in conceptual modeling will diagnose the anomalies more accurately than novices. | Not supported |

## Discussion

Clarifying the nature of expertise in conceptual modeling is crucial in order to improve business analysts' training curriculum. Even if the characteristics of visual attention related to optimized searches within visual stimuli are known, expertise can translate into several behavioral dimensions that - our results suggest - may be orthogonal, depending on the domain. Expert radiologists can identify an anomaly more quickly in a visual stimulus (Krupinski, 2000), while an expert probation officer will tend to take more time than a novice during a file reconstruction exercise (Lurigio et Carroll, 1985).

Thus, by comparing the performance, the visual attentional characteristics and the antecedents between business analysts and novices in conceptual modeling, our main objective was to deepen insight into the interdependent heuristics of experts and how dimensions of expertise affect the behavior and performance of the participants.

First, we compared the attentional characteristics of correct and incorrect diagnosis, in order to compare our results with prior experiments. While H1 was supported, indicating that correct diagnostics tend to take a shorter amount of time than incorrect answers, which concur with prior findings among the literature (Van Waes et al., 2009), H2 and H3 were not. Contrary to what has been observed in prior work, the accurate diagnostics in this experiment weren't linked with a shorter amount of fixation count or fixation duration (Van Waes et al., 2009; Henderson and Hollingworth, 1999; Holmqvist et al., 2011). While not statistically significant, the results are paradoxical and thus interesting, since they point toward an increase rather than a decrease in fixation count and duration. This might be explained by the complexity of the task. Indeed, it has been postulated that a lower amount of fixation could mean that the task was merely too simple, therefore necessitating a lesser amount of cognitive processing (Gegenfurtner et al. 2011). In contrast, the fact that the task may require a larger amount of fixation and fixation duration on a stimulus to be successfully completed may show that the anomaly detection process in conceptual modeling is more cognitively complex than similar tasks in other domains. This postulation is supported by known ambiguity of conceptual models and the challenge they present in terms of comprehension (Figl, 2017). It therefore becomes clear that a certain amount of fixation is needed to fully understand the models and that trials with a lesser amount of fixation count and duration are linked with incorrect answers, since the participants may have under-appreciated the conceptual richness of the stimulus, leading to premature and inaccurate diagnosis.

Second, by comparing the error detection process between the different error types, we aimed to better understand the relation between model comprehension and the semantic and syntactic dimensions of conceptual modeling. Syntactic errors involving invalid symbols use is the type of error diagnosed the fastest, leading us to surmise that they may be the easiest kind of error to spot. This concurs with prior work on syntactic and semantic errors, where syntactic errors are found to be less subtle and easier to recognize than semantic errors, leading to a lesser amount of time needed to be diagnosed (Davis et al., 2018; Dijkman et al., 2008). An interesting finding arises from the comparison of the two kinds of syntactic errors. Stimuli with seeded syntactic errors of non-consistent flow tend to be viewed for a longer amount of time than models with syntactic errors that used invalid symbols, thus suggesting greater complexity. While more tests and analysis are needed to better understand the full nature of this complexity, we are aware of no study that compared those two kinds of syntactic errors, and, therefore, we believe it is a lead worth investigating further. Contrary to expectations, stimuli with no seed errors were answered slower than other error types. This can be explained by the concept of 'stopping rule', which is the extent to which participants would continue or terminate their search for additional information before taking a decision (Browne and Pitts, 2004; Nickels et al., 1995). It is then reasonable to expect that the participant takes more time to diagnose that a model has no errors than the average time needed to diagnose a semantic or syntactic error, since they must pass through the same cognitive process, without stopping their search at the first error found.

The most interesting - and surprising - results appear when we compare experts with novices. Contrary to our hypotheses and prior studies, we found that experts were not more efficient and effective than novices.

We found that experts did not have fewer fixations, fixation duration or total view time than novices, and even had significantly lower performance than novices for error-free stimuli.

While some studies have found that experts may behave similarly to novices, especially in tasks requiring judgment (Bédard, 1989; Goldberg, 1959; Levy and Ulman, 1967), few experiments have produced results where novices performed better than experts (Adelson, 1984; Köpke and Nespoulous, 2006). In these experiments, the qualitative difference between how experts and novices perform the tasks would influence which task was more suited for novices and which for experts. The task type and complexity would then influence which group had a better performance. For example, Adelson (1984) found out that the type of representations constructed by computer programmers was different between novices and experts, and that each type of representation was more suited for a type of task.

This leads us to propose that our task type, or the level of complexity of the task, may have been more suited to novices: this might partly explain the unexpected results. Conceptual models are - by their very name and nature - potent with ambiguities that business analysts must process cognitively in order to find anomalies and errors. However, the use of simple models, as used in this experiment, could lead the experts to over-complexify the stimuli when trying to find all the ambiguities. Our finding that experts tend to diagnose more non-existent anomalies than novices, could lead to more false diagnoses, more time spent on the stimuli and more fixations.

Biases linked to expertise also need to be taken into consideration, since cognitive bias is considered one of the most serious handicaps of experts (Ericsson et al., 2006). Studies in physics have found that experts tend to activate the areas of the brain associated with inhibition more than novices, suggesting that experts have to inhibit misconceptions in order to give a sound answer (Brault Foisy et al., 2015; Masson et al., 2014; Masson, 2007). Furthermore, experts tend to have more design fixations, such as functional fixedness (i.e. restricting the use of an object to previously encountered functions) or mental sets, which limit the creativeness and set of ideas used in problems solving (Ericsson et al., 2006; Jansson and Smith, 1991). In our situation, this bias is extremely important, since understanding and diagnosing unknown models, without any context or clues, requires a fair amount of creativity and cognitive flexibility. Cognitive schemata arise from prior experiences. They frame and, in some ways, limit the business analysts ability to inhibit misconceptions. The creative limitations from this bias could very well explain why most experts tend to look extensively for an error and find ambiguities even in error-free stimuli. This claim is supported by the *post-hoc* analysis showing that high performing experts seem to manage to overcome this bias and successfully diagnose error-free stimuli. Clearly, further studies and experiments are warranted to explore the deeper insights indicated by the contradictory results.

Our main theoretical and practical contribution lay in the elaboration of the expert-novice dichotomy in conceptual modeling. Our findings regarding experts' performance, and the unexpectedly narrow difference between that of experts and novices, reveals a paradox that offers a new perspective on the richness of the cognitive *milieu* of expertise in modeling. From a theoretical perspective, our results highlight a clear need for more in-depth studies on how business analysts process and comprehend conceptual models compared to novices. At a practical level, articulating the difference between semantic and syntactic errors, and the difference between the two kinds of syntactic errors, will enable instructors to adapt and improve their curriculum when training new business analysts. Demonstration that different error types require different levels of cognitive processing, and that experts may over-complexify their representation of a model when looking for anomalies in error-free stimuli, are important to take in consideration when trying to develop novices into experts.

## Limitations and Future Research

This paper presents an initial attempt to articulate the differences between experts and novices in an anomaly detection task in conceptual modeling: the cognitive complexity of both the field and the study give rise to limitations and opportunities for improvement in future work.

While we use the concept of IT competence (Basselier et al., 2003) to propose a link between experience as a business analyst and expertise in conceptual modeling, we do not actually control or measure the level of IT competence of our participants. Rather than using a conceptual model based on the model of Basselier et al. (2003) to measure the antecedents of IT competence, we used years of experience as a business analyst as a surrogate indicator of competence. An interesting way to extend this experiment would be to study the actual independent variables of IT competence that would contribute to improving the business analysts' expertise in conceptual modeling. Furthermore, it would be interesting to widen the range or scope of expertise, by comparing participants in 3 or 4 groups with different levels of experience. This would allow us to further explore the orthogonality of the characteristics and 'dimensions' of expertise, rather than trying to compare what we now primarily see as a binary concept with two samples.

A further limitation of our study is the fact that we didn't control for domain expertise (or semantic expertise), and rather used simple and well-known scenarios. By collecting data on the familiarity of participants for each scenario, we may have a better insight into what really causes such small difference between novices and experts.

New measures, such as analysing and comparing the scanpath of novices and experts, have the potential to offer us valuable information on reading techniques used by experienced modelers which could, in turn, be used to provide recommendations for teaching curricula, the design of instructional materials and the revision of standardized notations such as BPMN.

Since the acquisition of skill can bring major changes to the brain activity and areas activated, depending on the skill and the training, a study using functional magnetic resonance imaging (fMRI) techniques or EEG might allow us to pinpoint the domain-general control areas and the domain-specific representational areas associated to expertise in error detection tasks during conceptual modelling. Considering that the performance of those areas is sensitive to the nature of the training, by identifying which areas of the brain to work on, we could create training curricula and materials strengthening those areas, thus improving the cognitive processing of future business analysts. Furthermore, a more detailed understanding of the cognitive processing of experts will grant us valuable insights into how novices and experts differ, overcoming the limitations of the apparently false dichotomy that currently persists in the literature.

## Acknowledgments

The authors are grateful for the constructive and helpful comments from the participants of the Vienna Retreat on NeuroIS 2018 in Vienna (Austria, June 19-21) and the participation of the members of the International Institute of Business Analysis (IIBA). This research was funded by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References

Adelson, B. When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 3 (1984), 483-495.

Bassellier, G.; Benbasat, I.; and Reich, B.H. The influence of business managers' IT competence on championing IT. *Information Systems Research*, 14, 4 (2003), 317-336.

Batra, D.; Hoffler, J.A.; and Bostrom, R.P. Comparing representations with relational and EER Models. *Communications of the ACM*, 33, 2 (1990), 126-139.

Bavota, G.; Gravino, C.; Oliveto, R.; De Lucia, A.; Tortora, G.; Genero, M.; and Cruz-Lemus, J.A. Identifying the Weaknesses of UML Class Diagrams during Data Model Comprehension. In: Whittle, J., Clark, T., Kühne, T. (eds.), *International Conference on Model Driven Engineering Languages and Systems (MODELS 2011)*, Heidelberg: Springer, 2011, pp. 168-182.

Becker, J.; Delfmann, P.; Dreiling, A.; Knackstedt, R.; and Kuropka, D. Configurative Process Modeling – Outlining an Approach to increased Business Process Model

Usability. In: Khosrow-Pour, M. (ed.), *Innovations Through Information Technology*, New Orleans: Gabler, 2004, pp. 615-619.

Bédard, J. Expertise in auditing: Myth or reality? *Accounting, Organizations and Society*, 14, 1-2 (1989), 113-131.

Bednarik, R., and Tukiainen, M. An eye-tracking methodology for characterizing program comprehension processes. In: Spencer, S.N. (ed.), *Proceedings of the 2006 symposium on Eye tracking research & applications*, New-York: ACM, 2006, pp. 125-132.

Bera, P. Does cognitive overload matter in understanding BPMN models? *Journal of Computer Information Systems*, 52, 4 (2012), 59–69.

Birkmeier, D.; Kloeckner, S.; and Overhage, S. An Empirical Comparison of the Usability of BPMN and UML Activity Diagrams for Business Users. In: Alexander, P.M., Turpin, M., van Deventer, J.P. (eds.), *18th European Conference on Information Systems*, Pretoria: ECIS, 2010, pp. 51-62.

Brault Foisy, L.-M.; Potvin, P.; Riopel, M.; and Masson, S. Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education*, 4, 1-2 (2015), 26-36.

Braze, D.; Shankweiler, D.; Ni, W.; and Palumbo, L.C. Readers' eye movements distinguish anomalies of form and content. *Journal of psycholinguistic research*, 31, 1 (2002), 25-44.

Browne, G.J., and Pitts, M.G. Stopping rule use during information search in design problems. *Organizational Behavior and Human Decision Processes*, 95, 2 (2004), 208-224.

Davis, C.J.; Hevner, A.R.; Labonte-LeMoyne, É.; and Léger, P.-M. Expertise as a Mediating Factor in Conceptual Modeling. In: Davis, F., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. (eds.), *Information Systems and Neuroscience*, Cham: Springer, 2018, pp. 85-92.

Davis, C.J., and Hufnagel, E.M. Through the eyes of experts: A socio-cognitive perspective on the automation of fingerprint work. *MIS Quarterly*, 31, 4 (2007), 681-703.

Dijkman, R.M.; Dumas, M.; and Ouyang, C. Semantics and analysis of business process models in BPMN. I*nformation and Software Technology*, 50, 12 (2008), 1281-1294.

Eikebrokk, T.R.; Iden, J.; Olsen, D.H.; and Opdahl, A.L.: Exploring Process-Modelling Practice: Towards a Conceptual Model. In: Sprague Jr., R.H. (ed.), *Proceedings of the 41st Annual Hawaii International Conference on System Sciences,* Waikoloa: IEEE, 2008, pp. 376-382.

Ericsson, K.A.; Charness, N.; Feltovich, P.J.; and Hoffman, R.R. *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge: Cambridge University Press, 2006.

Figl, K. Comprehension of procedural visual business process models. *Business & Information Systems Engineering,* 59, 1 (2017), 41-67.

Figl, K., and Strembeck, M. Findings from an Experiment on Flow Direction of Business Process Models. In: Kolb, J., Leopold, H., Mendling, J. (eds.), *Proceedings of the 6th International Workshop on Enterprise Modelling and Information Systems Architectures*, Bonn: Koellen, 2015, pp. 59-73

Figl, K.; Mendling, J.; and Strembeck, M. The influence of notational deficiencies on process model comprehension. J*ournal of the Association for Information Systems*, 14, 6 (2013), 312–338.

Gegenfurtner, A.; Lehtinen, E.; and Säljö, R. Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23, 4 (2011), 523-552.

Gemino, A., and Wand, Y. A framework for empirical evaluation of conceptual modeling techniques. *Requirements Engineering*, 9, 4 (2004), 248-260.

Glaser, R. Education and thinking: The role of knowledge. *American psychologist*, 39, 2 (1984), 93-104.

Gobet, F. *Understanding Expertise: A Multi-Disciplinary Approach*. London: Red Globe Press, 2016.

Goldberg, L. R. The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. *Journal of Consulting Psychology,* 23, 1 (2959), 25-33.

Goolkasian, P. Pictures, words, and sounds: From which format are we best able to reason? *The Journal of General Psychology*, 127, 4 (2000), 439-459.

Henderson, J.M., and Hollingworth, A. The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10, 5 (1999), 438-443.

Herbig, B., and Büssing, A. The role of explicit and implicit knowledge in work performance. *Psychology Science*, 46, 4 (2004), 408-432.

Hill, N. M., and Schneider, W. (2006). Brain Changes in the Development of Expertise: Neuroanatomical and Neurophysiological Evidence about Skill-Based Adaptations. In: Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (eds.), *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge: Cambridge University Press, 2006, pp. 653-682.

Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; and Van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures,* Oxford: Oxford University Press, 2011.

Indulska, M.; Recker, J.; Rosemann, M.; and Green, P. Business Process Modeling: Current Issues and Future Challenges. In: van Eck, P., Gordijn, J., Wieringa, R. (eds.), *21st International Conference on Advanced Information Systems Engineering*, Berlin: Springer, 2009, pp. 501-514.

Jabbari Sabegh, M. A.; Lukyanenko, R.; Recker, J. C.; Samuel, B.; and Castellanos, A. Conceptual Modeling Research in Information Systems: What we now know and what we still do not know. In: Samuel, B., Lukyanenko, R., Castellanos, A. (eds.), *AIS Symposium on Research in Systems Analysis and Design (AIS SIGSAND 2017)*, Cincinnati: Springer, 2017, pp. 14-23.

Jansson, D. G., and Smith, S. M. Design fixation. *Design studies*, 12, 1 (1991), 3-11.

Jošt, G.; Huber, J.; Heričko, M.; and Polančič, G. An empirical investigation of intuitive understandability of process diagrams. *Computer Standards & Interfaces*, 48, (2016), 90-111.

Just, M.A., and Carpenter, P.A. Eye fixations and cognitive processes. *Cognitive psychology*, 8, 4 (1976), 441-480.

Köpke, B., and Nespoulous, J. L. Working memory performance in expert and novice interpreters. *Interpreting*, 8, 1 (2006), 1-23.

Koschmider, A.; Ullrich, M.; Heine, A.; and Oberweis, A. Revising the Vocabulary of Business Process Element Labels. In: Zdravkovic, J., Kirikova, M., Johannesson, P. (eds.), *International Conference on Advanced Information Systems Engineering*, Cham: Springer, 2015, pp. 69–83.

Kummer, T.F.; Recker, J.; and Mendling, J. Enhancing understandability of process models through cultural-dependent color adjustments. *Decision Support Systems*, 87, (2016), 1-12.

Krupinski, E.A. The importance of perception research in medical imaging. *Radiation medicine*, 18, 6 (2000), 329-334.

Kung, C. H., and Solvberg, A. Activity Modeling and Behaviour Modeling. In: Olle, T.W., Sol, H.G., Verrijn-Stuart, A.A. (eds.), *Proceedings of the IFIP WG 8.1 working conference on Information System Design Methodologies: Improving the Practice*, Noordwijkerhout: North-Holland Publishing Co., 1986, pp. 145–171.

Larkin, J.H., and Simon, H. A. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11, 1 (1987), 65-100.

Levy, B. I., and Ulman, E. Judging psychopathology from paintings. *Journal of abnormal psychology*, 72, 2 (1967), 182-187.

Liberatore, M.J.; Hatchuel, A.; Weil, B.; and Stylianou, A.C. An organizational change perspective on the value of modeling. *European Journal of Operational Research*, 125, 1 (2000), 184-194.

Lupu, R. G., and Ungureanu, F. A survey of eye tracking methods and applications. *Buletinul Institutului Politehnic din Iasi, Automatic Control and Computer Science Section*, 3, (2013), 72-86.

Lurigio, A.J., and Carroll, J.S. Probation officers' schemata of offenders: Content, development, and impact on treatment decisions. *Journal of Personality and Social Psychology*, 48, 5 (1985), 1112-1126.

Masson, S.  Enseigner les sciences en s'appuyant sur la neurodidactique des sciences. In: Potvin, P., Riopel, M., Masson, S. (eds.), *Enseigner les Sciences: Regards Multiples*, Québec: Éditions MultiMondes, 2007, pp. 308-321

Masson, S.; Potvin, P.; Riopel, M.; and Foisy, L. M. B. Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, 8, 1 (2014), 44-55.

Mendling, J.; Reijers, H.A.; and Recker, J. Activity labeling in process modeling: Empirical insights and recommendations. *Information Systems*, 35, 4 (2010), 467-482.

Mendling, J., and Strembeck, M. Influence Factors of Understanding Business Process Models. In: Abramowicz, W., Fensel, D. (eds.), *International Conference on Business Information Systems*, Heidelberg: Springer, 2008, pp. 142–153.

Moody, D. The “physics” of notations: toward a scientific basis for constructing visual notations in software engineering. *IEEE Transactions on Software Engineering,* 35, 6 (2009), 756-779.

Ni, W.; Fodor, J. D.; Crain, S.; and Shankweiler, D. Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, 27, 5 (1998), 515-539.

Nickles, K.R.; Curley, S.P.; and Benson, P.G. Judgment-based and reasoning-based stopping rules in decision making under uncertainty. Working Paper, Wake Forest University, (1995), 1-38.

Ottensooser, A.; Fekete, A.; Reijers, H.A.; Mendling, J.; and Menictas, C. Making sense of business process descriptions: an experimental comparison of graphical and textual notations. *Journal of Systems and Software*, 85, 3 (2012), 596–606.

Parsons, J., and Cole, L. What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques. *Data & Knowledge Engineering*, 55, 3 (2005), 327-342.

Patel, V. L., and Groen, G. J. The General and Specific Nature of Medical Expertise: A Critical Look. In: Ericsson, K.A., Smith, J. (eds.), *Toward a General Theory of Expertise: Prospects and Limits*, Cambridge: Cambridge University Press, 1991, pp. 93-125.

Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124, 3 (1998), 372-422.

Recker, J. Empirical investigation of the usefulness of gateway constructs in process models. *European Journal of Information Systems,* 22, 6 (2013), 673–689

Recker, J.C., and Dreiling, A. Does it Matter which Process Modelling Language We Teach or Use? An Experimental Study on Understanding Process Modelling Languages Without Formal Education. In: Tan, W.-G. (ed.), *Proceedings of the ACIS 2007 18th Australasian Conference on Information Systems Doctoral Consortium*, Queensland: University of Southern Queensland, 2007, pp. 356-366.

Recker, J.; Reijers, H.A.; and van de Wouw, S.G. Process model comprehension: the effects of cognitive abilities, learning style, and strategy. *Communications of the Association for Information Systems*, 34, 9 (2014), 199–222.

Recker, J.; Safrudin, N.; and Rosemann, M. How novices design business processes. *Information Systems*, 37, 6 (2012), 557-573.

Reijers, H.A., and Mendling, J. A study into the factors that influence the understandability of business process models. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41, 3 (2011), 449–462.

Reingold, E.M., and Sheridan, H. Eye Movements and Visual Expertise in Chess and Medicine. Oxford handbook on eye movements. In: Liversedge, S.P., Gilchrist, I.D., and

Everling, S. (eds.), *Oxford Handbook on Eye Movements*, Oxford: Oxford University Press, 2011, pp. 528-550.

De A. Rodrigues, A.; De O. Barros, M..; Revoredo, K.; Azevedo, L.G.; and Leopold, H. An Experiment on Process Model Understandability Using Textual Work Instructions and BPMN Models. In: O'Conner, L. (ed.), 2*9th Brazilian Symposium on Software Engineering (SBES)*, Belo Horizonte: IEEE, 2015, pp. 41–50.

Sánchez-González, L.; García, F.; Mendling, J.; and Ruiz, F. Quality Assessment of Business Process Models Based on Thresholds. In: Meersman, R., Dillon, T., Herrero, P. (eds.), *On the Move to Meaningful Internet Systems: OTM 2010*, Heidelberg: Springer, 2010, pp. 78–95.

Sandkuhl, K., and Wiebring, J. Experiences from Selecting a BPM Notation for an Enterprise. In: Abramowicz, W. (ed.), *Business Information Systems Workshops (BIS 2015)*, Cham: Springer, 2015, pp. 126-138.

Shanks, G. Conceptual data modelling: an empirical study of expert and novice data modellers. *Australasian Journal of Information Systems*, 4, 2 (1997), 63-73.

Sheridan, H., and Reingold, E.M. Expert vs. novice differences in the detection of relevant information during a chess game: evidence from eye movements. *Frontiers in Psychology*, 5, (2014), 941-946.

Sonnentag, S. Expertise at work: Experience and excellent performance. In: Cooper, C.L., Robertson, I.T. (eds.), *International Review of Industrial and Organisational Psychology*, Chichester: Wiley, 2000, pp. 223-264.

Speelman, C. Implicit Expertise: Do We Expect Too Much From our Experts? In: Kirsner, K., Speelman, C., Maybery, M., O´Brien-Malone, A., Anderson, M., MacLeod, C. (eds.), *Implicit and Explicit Mental Processes*, Mawah: Erlbaum, 1998, pp. 135-147.

Turetken, O.; Rompen, T.; Vanderfeesten, I.; Dikici, A.; and van Moll, J. The Effect of Modularity Representation and Presentation Medium on the Understandability of Business Process Models in BPMN. In: La Rosa, M., Loos, P., Pastor, O. (eds.), *Business Process Management (BPM 2016)*, Cham: Springer, 2016, pp. 289–307.

Van Waes, L.; Leijten, M.; and Quinlan, T. Reading during sentence composing and error correction: A multilevel analysis of the influences of task complexity. *Reading and Writing,* 23, 7 (2009), 803-834.

Wahl, T., and Sindre, G. An Analytical Evaluation of BPMN Using a Semiotic Quality Framework. In: Siau, K. (ed.), *Advanced Topics in Database Research*, Hershey: IGI Global, 2006, pp. 94-105.

Wand, Y., and Weber, R. Research commentary: Information systems and conceptual modeling—A research agenda. *Information Systems Research*, 13, 4 (2002), 363-376.

Weitlaner, D.; Guettinger, A.; and Kohlbacher, M. Intuitive Comprehensibility of Process Models. In: Fischer, H., Schneeberger, J. (eds.), *S-BPM ONE—Running Processes*, Heidelberg: Springer, 2013, pp. 52–71.

Wineburg, S. Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science*, 22, 3 (1998), 319-346.

Yusuf, S.; Kagdi, H.; and Maletic, J.I. Assessing the Comprehension of UML Class Diagrams via Eye Tracking. In: Wong, K., Stroulia, E., Tonella, P. (eds.), *15th IEEE International Conference on Program Comprehension,* Banff: IEEE, 2007, pp. 113-122.

Zhan, Z.; Zhang, L.; Mei, H.; and Fong, P.S. Online learners' reading ability detection based on eye-tracking sensors. *Sensors*, 16, 9 (2016) 1457-1473.

## Appendix

### 1) Questionnaire on past experience with conceptual modeling (translated from French)

Q1: Select the graphical conceptual modeling notation(s) that you have used in the past.

▢ Business Process Model and Notation - BPMN

▢ Unified Modelling Language - UML

▢ Event-driven process chain - EPC

▢ Entity-relationship model - ER

▢ Extended Enterprise Modeling Language - EEML

▢ Petri net

▢ Other (please specify)

_______________________________________________

▢ I have never used any graphical notation in the past

Q2: On how many projects involving the use or creation of conceptual models have you worked?

o 0 project
o 1 to 3 projects
o 4 to 6 projects
o 7 to 10 projects
o More than 10 projects

Q3: How many hours did you spend working on conceptual models?

o 0 - 15 hours
o 16 - 50 hours
o 51 - 100 hours
o 101 - 150 hours
o 151 - 200 hours
o More than 200 hours

Q4: What interaction(s) do you normally have with conceptual models?

- ▢ Read the models
- ▢ Design the models
- ▢ Check the quality of the models
- ▢ Other (please specify)

____________________________________________

- ▢ No interaction

**The questions in the following section will deal with the Business Process Model and Notation (BPMN).**

Q5: On how many projects involving the use or creation of BPMN models have you worked?

- o 0 project
- o 1 to 3 projects
- o 4 to 6 projects
- o 7 to 10 projects
- o More than 10 projects

Q6: How many hours did you spend working on BPMN models?

- o 0 - 15 hours
- o 16 - 50 hours
- o 51 - 100 hours
- o 101 - 150 hours
- o 151 - 200 hours
- o More than 200 hours

Q7: What interaction(s) do you normally have with BPMN models?

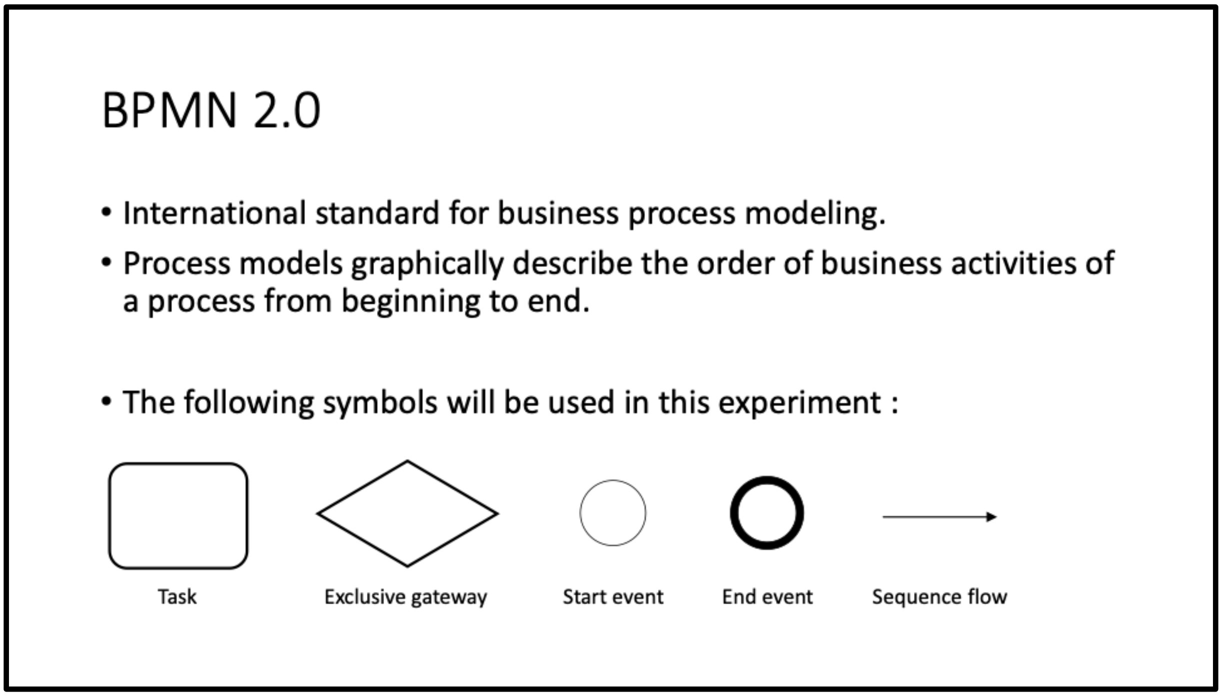
- ▢ Read the models
- ▢ Design the models

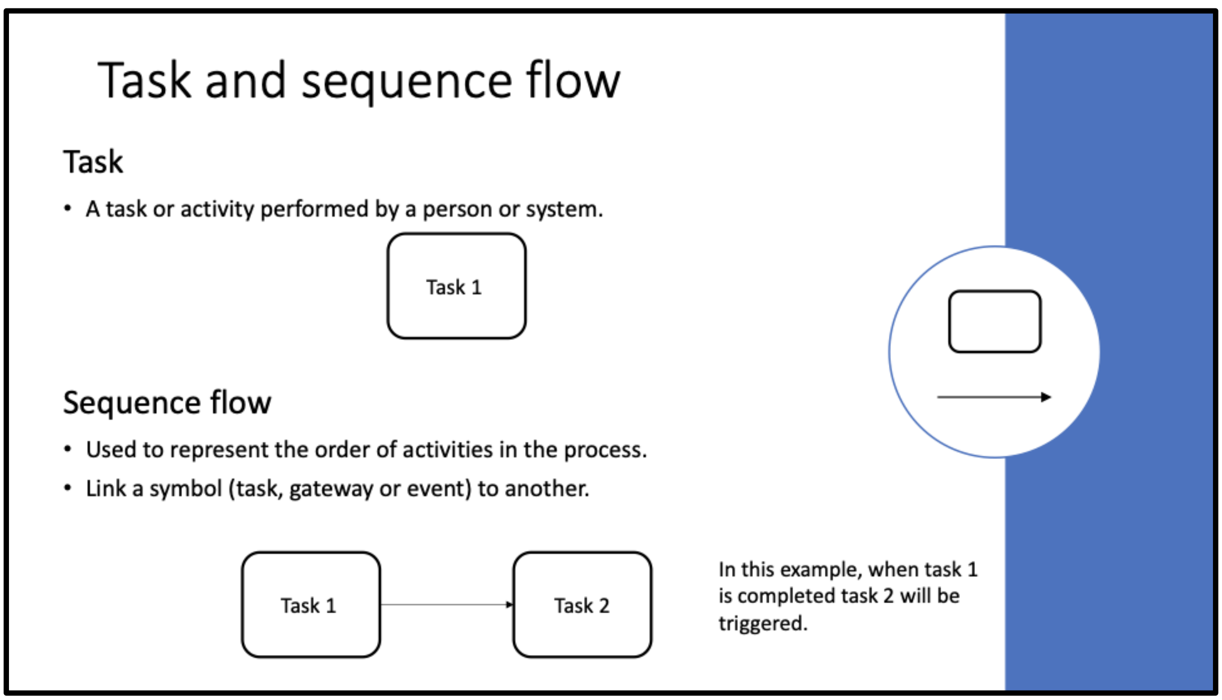
- ☐ Check the quality of the models
- ☐ Other (please specify)

_______________________________________________

- ☐ No interaction

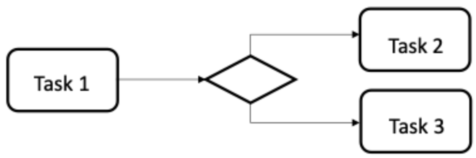
## 2) Presentation on BPMN (translated from French)



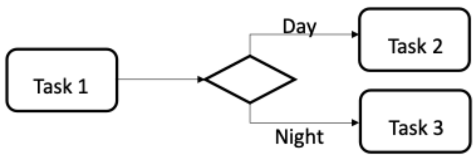
## Exclusive gateway

- When splitting, the gateway guides the flow to exactly one of the outgoing branches, based on conditions.




In this example, the gateway will redirect the flow to **either** task 2 or task 3, depending on conditions or a decision.

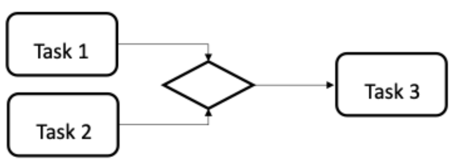
- Conditions can be specified on the sequence flow.




In this example, the gateway will redirect the flow to **either** task 2 or task 3, depending on the time of day.

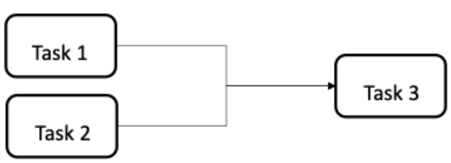
## Exclusive gateway

- When merging, the gateway waits for an incoming branch to trigger the outgoing flow.




In this example, when the flow from task 1 **or** the flow from task 2 reaches the gateway, the outbound flow to task 3 will be triggered.

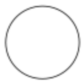
- This notation is however **optional**.



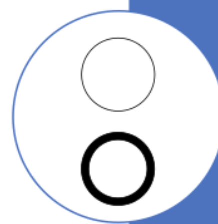

These two notations are **equivalent**.

# Events

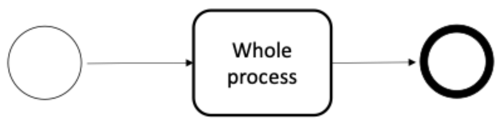
- Mark the beginning of the process.



- Mark the end of the process.



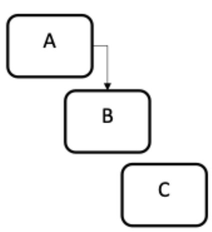
- For a model to be valid, every possible path must lead to the end of the process. We must therefore avoid "dead ends".



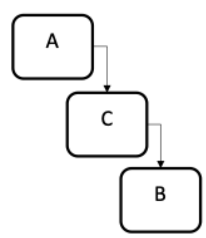

# Types of errors

- During the experiment, you will be led to identify errors in conceptual models and categorize them into two groups:

**Syntactic errors**

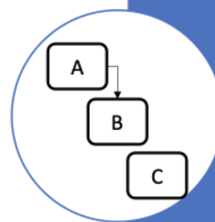



**Semantic errors**

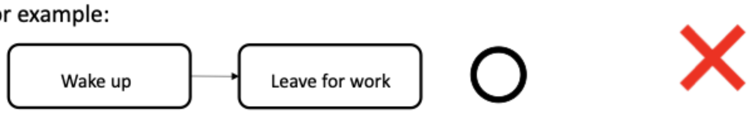


# Syntactic error

- How different symbols can be combined with each other.
- Composed of a symbolic vocabulary and grammar.
- A syntactic error is a misuse of symbols or relationships between different symbols.




For example:

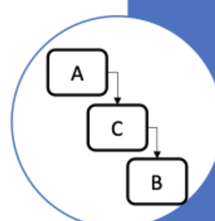



The link between the tasks and the end of the process is missing, making it impossible to complete the process. Therefore, this is a syntactic error.

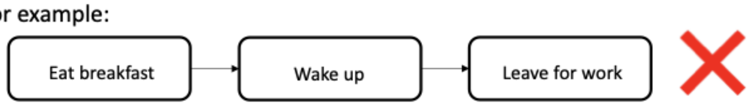
The corrected model would be the following:


# Semantic error

- Semantic is the element that gives meaning to each symbol and the relations between symbols.
- Represents the diagram message.
- A semantic error is a logic error in the process.




For example:




It down not make sense to have breakfast before waking up Therefore, this is a semantic error.

The corrected model would be the following :

# Chapitre 4 : Conclusion

L'objectif principal de ce mémoire était d'identifier les « meilleures pratiques » et les éléments les plus complexes de la modélisation conceptuelle et d'approfondir notre compréhension des caractéristiques de l'attention visuelle dans une tâche de détection et de diagnostic d'erreurs dans des modèles de processus d'affaires. Plus précisément, ce mémoire cherche à poser des recommandations sur le développement et l'amélioration des programmes de formation en analyse d'affaires.

Une expérience en laboratoire intra-sujet a été réalisée à l'hiver 2018 afin de tester nos hypothèses et répondre à notre question de recherche. 30 participants ont été recrutés et ont été rémunérés à l'aide d'un coupon COOP HEC de 20$. Ceux-ci devaient identifier et diagnostiquer des erreurs dans des modèles conceptuels en BPMN. Les caractéristiques de l'attention visuelle des participants lors de la recherche d'erreurs sur les stimuli ont été collectées par un oculomètre et l'expérience passée des participants a été mesurée à l'aide d'un questionnaire. Cette collecte de données a permis à l'étudiant de ce mémoire de rédiger deux articles complémentaires.

Ce dernier chapitre fait un rappel à la question de recherche de ce mémoire et présente les principaux résultats des deux articles. Les contributions, tant théoriques que pratiques, ainsi que les limites de cette étude et les recommandations pour des études futures sont finalement soulevées.

## 4.1 Rappel de la question de recherche et principaux résultats

Les résultats de ces deux articles ont permis de répondre à la question de recherche de ce mémoire, soit :

*Quelles sont les différences relatives à l'attention visuelle entre les analystes d'affaires experts et les novices, lors d'une tâche de détection d'erreurs dans des modèles conceptuels ?*

Trois hypothèses ont été formulées dans le premier article et quatre hypothèses ont été rajoutées dans le deuxième article, en se basant sur la littérature de la modélisation conceptuelle, l'expertise et l'attention visuelle, afin de répondre à cette question de recherche.

**H1 : Les détections d'erreur réussies en modélisation conceptuelle demanderont moins de temps de vue du stimulus que les détections d'erreur infructueuses. (Supportée)**

En effet, les deux articles ont confirmé que les diagnostics exacts, où les participants ont identifié la bonne erreur et l'ont associée au bon type d'erreur, étaient répondus plus rapidement que les diagnostics erronés. De plus, les modèles avec des erreurs syntaxiques de mauvaise utilisation de symboles (ErrSyn2) étaient répondus plus rapidement que les modèles avec des erreurs syntaxiques de flux incohérent (ErrSyn3). Les stimuli sans erreurs étaient ceux qui étaient regardés le plus longtemps avant d'être diagnostiqués.

Ces résultats concordent avec la littérature et indiquent que plus un participant passe du temps sur un même stimulus, moins il y a de chances d'identifier correctement l'erreur, car trop de ressources cognitives sont utilisées pour simplement encoder et comprendre le stimulus (Van Waes et coll. 2009).

**H2 : Les détections d'erreur réussies en modélisation conceptuelle nécessiteront, au total, moins de fixations que les détections d'erreur infructueuses. (Non supportée)**

Les analyses préliminaires du premier article démontrent que les diagnostics exacts sont liés à une baisse du nombre de fixations et une augmentation de la proportion des fixations sur les zones d'intérêt (c.-à-d. les zones contenant les erreurs), autant pour les erreurs sémantiques que syntaxiques.

Toutefois, l'analyse complète des résultats, tirée du deuxième article, contredit les résultats préliminaires du premier article. En effet, les bons diagnostics sont liés, à peine significativement, à une hausse du nombre de fixations. Par contre, tout comme le premier article, la proportion des fixations sur les zones d'intérêt était plus élevée lors des bons diagnostics, autant pour les erreurs sémantiques que syntaxiques.

**H3 : Les détections d'erreur réussies en modélisation conceptuelle nécessiteront, en moyenne, une durée de fixation plus courte que les détections d'erreur infructueuses. (Non supportée)**

Lors des analyses préliminaires du premier article, un lien significatif a été identifié entre les bons diagnostics et une baisse de la durée moyenne des fixations.

Toutefois, lors de l'analyse finale du deuxième article, aucun résultat significatif ne fut relevé.

Contrairement à ce que la littérature suggère, les bons diagnostics n'étaient pas liés à une durée de fixation moyenne plus courte que les mauvaises réponses (Van Waes et coll. 2009). Ce résultat peut être expliqué par l'ambiguïté inhérente aux modèles conceptuels ; un certain niveau d'effort est nécessaire pour bien comprendre les modèles et les participants qui ont répondu trop rapidement avaient peut-être sous-estimé la richesse des stimuli, menant à un diagnostic prématuré et inexact (Figl, 2017).

**H4 : Les experts en modélisation conceptuelle nécessiteront moins de temps que les novices pour examiner le stimulus. (Non supportée)**

Aucune relation significative ne fut observée entre l'expertise et le temps de réponse. Ainsi, les experts n'ont pas répondu plus rapidement ou plus lentement que les novices. Une analyse plus poussée du comportement des experts nous a permis de remarquer qu'ils répondaient plus rapidement aux modèles ayant des erreurs syntaxiques de mauvaise utilisation de symboles, et moins rapidement aux stimuli n'ayant aucune erreur.

**H5 : Les experts en modélisation conceptuelle nécessiteront, au total, moins de fixations que les novices. (Non supportée)**

Encore une fois, aucun lien significatif n'a été observé entre l'expertise et le nombre de fixations.

**H6 : Les experts en modélisation conceptuelle nécessiteront, en moyenne, d'une durée de fixation plus courte que celle des novices. (Non supportée)**

Tout comme H5, aucun lien significatif ne peut être tiré entre l'expertise et la durée moyenne de fixation.

**H7 : Les experts en modélisation conceptuelle diagnostiqueront plus précisément les anomalies que les novices. (Non supportée)**

Contrairement à l'hypothèse, les experts n'avaient pas, de façon significative, une meilleure performance que les novices lors de l'identification et du diagnostic de modèles ayant des erreurs sémantiques ou syntaxiques. Contradictoirement, les experts avaient tendance, de façon significative, à avoir un moins bon score que les novices pour le diagnostic de modèle sans erreur.

Les résultats aux hypothèses H4, H5, H6 et H7 sont surprenants et révèlent que les experts ne semblent pas être plus efficaces ni précis dans leur recherche d'erreurs dans les modèles conceptuels. Au contraire, nous notons que les experts avaient une performance inférieure à celle des novices pour les stimuli n'ayant aucune erreur. Plusieurs raisons peuvent expliquer ces résultats, telles que la complexité de la tâche ou des stimuli plus appropriés pour les novices ou un biais lié à l'expertise.

Un biais d'expertise tel que la fixité fonctionnelle ou de conception, qui est le fait de restreindre l'utilisation d'un objet ou d'un élément aux fonctions déjà rencontrées, peut grandement limiter la créativité d'un participant, et donc sa performance dans notre étude (Jansson et Smith, 1991). Les limitations créatives de ce biais pourraient très bien expliquer pourquoi la plupart des experts ont tendance à rechercher une erreur et à trouver des ambiguïtés même dans des stimuli sans erreur. Cette analyse est corroborée par l'analyse post-hoc montrant que les participants du sous-groupe d'experts performants semblaient avoir réussi à surmonter ce biais et à diagnostiquer avec succès des stimuli sans erreur.

## 4.2 Contributions

D'un point de vue théorique, les résultats présentés dans ce mémoire contribuent à combler les lacunes dans la littérature concernant la nature et les caractéristiques de l'expertise en modélisation conceptuelle. Il est essentiel de comprendre comment les experts analysent et cherchent des erreurs dans des modèles conceptuels, afin de faciliter l'apprentissage de ces techniques aux futurs analystes d'affaires. Nos résultats démontrent que les experts ont tendance à identifier des ambiguïtés dans des modèles sans erreurs, et qu'ils ne sont pas particulièrement plus performants que les novices lorsqu'ils doivent trouver des erreurs dans des modèles sans contexte. Ces résultats sont particulièrement intéressants et ouvrent la porte vers des études plus précises sur les biais d'expertise en modélisation conceptuelle et sur la différence entre les connaissances du domaine (sémantique) et de la notation visuelle (syntaxique).

De plus, en comparant les processus d'identification d'erreur sémantique et syntaxique, nous participons à la littérature relativement peu développée sur l'effet des règles syntaxiques sur la compréhension des modèles (Moody, 2009). Nos résultats finaux sont en accord avec la littérature qui prétend que les erreurs sémantiques sont plus subtiles, et donc difficile à identifier (Davis et coll., 2018 ; Dijkman et coll., 2008). Par contre, en étant l'une des premières études comparant les caractéristiques de l'attention visuelle entre l'identification d'erreur syntaxique de mauvaise utilisation de symbole et de flux incohérent, nous établissons la base théorique pour de prochaines recherches.

Les implications pratiques de ce mémoire se traduisent en recommandations pour améliorer les curriculums et méthodes de formation des futurs analystes d'affaires. En effet, il est très pertinent pour les instructeurs et gestionnaires de comprendre les lacunes des analystes d'affaires lors de l'utilisation et vérification de modèles conceptuels. Être conscient des biais potentiels liés à l'expertise leur permet d'adapter leur curriculum en portant, par exemple, plus d'attention sur la créativité des analystes d'affaires en augmentant le nombre de tâches de compréhension et d'identification d'erreurs dans des modèles sans contexte. Finalement, le recours à des exercices poussant les analystes d'affaires à modeler le même processus d'affaires de différentes manières pourrait, encore une fois, contribuer au développement de leur créativité et venir diminuer leur biais de fixité de conception dans le futur.

## 4.3 Limites de l'étude et pistes de recherche future

Ce mémoire présente une première étude essayant d'explorer la différence entre experts et novices dans une tâche de détection et de diagnostic d'anomalie en modélisation conceptuelle. Il présente donc certaines limites et possibilités d'amélioration dans le futur.

Plutôt que d'utiliser un modèle conceptuel basé sur le modèle de Basselier et coll. (2003) pour mesurer les antécédents de compétences TI, nous avons utilisé le nombre d'années d'expérience en tant qu'analystes d'affaires comme indicateur de compétence. Une façon intéressante d'étendre cette expérience serait d'élargir l'éventail des compétences en comparant des participants de 3 ou 4 groupes ayant différents niveaux d'expérience. Cela permettrait d'étudier en profondeur l'échelle d'expertise, plutôt que d'essayer de comparer un concept non binaire avec deux échantillons.

L'un des principaux inconvénients de notre étude est le fait que nous n'avons pas tenu compte de l'expertise de domaine et avons plutôt limité son effet en utilisant des scénarios simples et généralement connus par tout le monde. En contrôlant la familiarité des participants pour chaque scénario, nous pouvons avoir une meilleure idée de ce qui cause réellement une si petite différence entre novices et experts.

De plus, de nouvelles mesures telles que l'analyse et la comparaison du parcours visuel des novices et des experts pourraient nous fournir des informations précieuses sur les techniques de lecture utilisées par les analystes d'affaires, qui pourraient à leur tour être utilisées pour faciliter et accélérer la formation de nouveaux analystes d'affaires.

Augmenter la taille de l'échantillon serait pertinent pour venir confirmer la validité de nos résultats. En effet, on remarque une certaine divergence entre les résultats des deux articles composant ce mémoire, même si les deux collectes de données étaient identiques. Cet écart peut potentiellement être expliqué par la différence dans la grosseur des échantillons utilisés lors des analyses statistiques, soit 18 participants pour le premier article et 30 participants pour le deuxième. Répliquer l'étude avec un échantillon plus grand pourrait alors augmenter la validité externe de nos résultats et venir renforcer nos conclusions.

Finalement, puisque l'acquisition et le développement de nouvelles compétences peuvent entraîner des modifications majeures de l'activité cérébrale et les zones activées, une étude utilisant la technique de l'imagerie par résonance magnétique fonctionnelle (IRMf) ou, dans une certaine mesure, l'électroencéphalographie (EEG), nous permettrait d'identifier les zones activées par l'expertise lors de la détection d'erreur dans la modélisation conceptuelle. Puisque la nature des tâches de formation influence la performance de ces zones, en les identifiant, nous pouvons développer des programmes de formation ciblés sur celles-ci, accélérant potentiellement l'acquisition de compétence. En outre, une meilleure compréhension du traitement cognitif des analystes d'affaires nous fournirait des informations précieuses sur les différences entre novices et experts. Toutefois, ces techniques sont plus coûteuses et, particulièrement pour l'IRMf, augmentent considérablement la complexité et les risques d’erreurs potentiels de l’expérience.

# Bibliographie

Aguilar-Savén, Ruth Sara (2004). « Business process modelling: Review and framework », International Journal of Production Economics, vol. 90, no. 2, p. 129-149.

Becker, Jörg, Patrick Delfmann, Alexander Dreiling, Ralf Knackstedt et Dominik Kuropka (2004). « Configurative process modeling–outlining an approach to increased business process model usability », dans *Innovations Through Information Technology*, New Orleans, Gabler, p. 615-619.

Boutin, Karl-David, Pierre-Majorique Léger, Christopher J. Davis, Alan R. Hevner et Élise Labonté-LeMoyne (2019). « Attentional Characteristics of Anomaly Detection in Conceptual Modeling », dans *Information Systems and Neuroscience*, Cham, Springer, p. 57-63.

Davis, Christopher J., Alan R. Hevner, Élise Labonté-LeMoyne et Pierre-Majorique Léger (2018). « Expertise as a Mediating Factor in Conceptual Modeling », dans *Information Systems and Neuroscience*, Cham, Springer, p. 85-92.

Dijkman, Remco M., Marlon Dumas et Chun Ouyang (2008). « Semantics and analysis of business process models in BPMN », Information and Software Technology, vol. 50, no. 12, p. 1281-1294.

Eikebrokk, Tom R., Jon Iden, Dag H. Olsen et Andreas L. Opdahl (2008). « Exploring Process-Modelling Practice: Towards a Conceptual Model», dans *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, Waikoloa, IEEE, p. 376-382.

Figl, Kathrin (2017). « Comprehension of procedural visual business process models », Business & Information Systems Engineering, vol. 59, no. 1, p. 41-67.

Gartner (2018). *Gartner Worldwide IT Spending Forecast*, Gartner. Récupéré le 20 novembre 2018 de https://www.gartner.com/technology/research/it-spending-forecast/

Indulska, Marta, Jan Recker, Michael Rosemann et Peter Green (2009). « Business Process Modeling: Current Issues and Future Challenges », dans *21st International Conference on Advanced Information Systems Engineering*, Berlin, Springer, p. 501-514.

Jabbari Sabegh, Mohammad Ali, Roman Lukyanenko, Jan C. Recker, Binny Samuel et Arturo Castellanos (2017). « Conceptual Modeling Research in Information Systems: What we now know and what we still do not know », dans *AIS Symposium on Research in Systems Analysis and Design (AIS SIGSAND 2017)*, Cincinnati, Springer, p. 14-23.

Jansson, David G. et Steven M. Smith (1991). « Design fixation », Design studies, vol. 12, no. 1, p. 3-11.

Jeston, John, et Johan Nelis (2010). *Business Process Management: Practical Guidelines to Successful Implementations*, London, Routledge, 500 p.

Kark, Khalid, Anjali Shaikh et Caroline Brown (2017). *Technology budgets: From value preservation to value creation*, Deloitte. Récupéré le 20 novembre 2018 de https://www2.deloitte.com/insights/us/en/focus/cio-insider-business-insights/technology-investments-value-creation.html

Moody, Daniel (2009). « The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering », IEEE Transactions on Software Engineering, vol. 35, no. 6, p. 756-779.

Ottensooser, Avner, Alan Fekete, Hajo A. Reijers, Jan Mendling et Con Menictas (2012). « Making sense of business process descriptions: an experimental comparison of graphical and textual notations », Journal of Systems and Software, vol. 85, no. 3, p. 596–606.

Recker, Jan, et Alexander Dreiling (2007). « Does it Matter which Process Modelling Language We Teach or Use? An Experimental Study on Understanding Process Modelling Languages Without Formal Education » dans *Proceedings of the ACIS 2007 18th Australasian Conference on Information Systems Doctoral Consortium*, Queensland, University of Southern Queensland, p. 356-366.

Recker, Jan, Michael Rosemann Marta Indulska, et Peter Green (2009). « Business process modeling-a comparative analysis », Journal of the Association for Information Systems, vol. 10, no. 4, p. 333-363.

Rodrigues, Raphael De A., Márcio De O. Barros, Kate Revoredo, Leonardo G. Azevedo et Henrik Leopold (2015). « An Experiment on Process Model Understandability Using Textual Work Instructions and BPMN Models » dans *29th Brazilian Symposium on Software Engineering (SBES)*, Belo Horizonte, IEEE, p. 41–50.

Van Waes, Luuk, Mariëlle Leijten et Thomas Quinlan (2009). « Reading during sentence composing and error correction: A multilevel analysis of the influences of task complexity », Reading and Writing, vol. 23, no. 7, p. 803-834.

Wand, Yair, et Ron Weber (2002). « Research commentary: Information systems and conceptual modeling—A research agenda », Information Systems Research, vol. 13, no. 4, p. 363-376.

Yusuf, Shehnaaz, Huzefa Kagdi et Jonathan I. Maletic (2007). « Assessing the Comprehension of UML Class Diagrams via Eye Tracking », dans *15th IEEE International Conference on Program Comprehension*, Banff, IEEE, p. 113-122.