



CASE STUDIES

HCI - P5 - WEEK 3

USER STUDIES

EMMA BEAUXIS-AUSSALET

e.m.a.l.beauxisaussalet@vu.nl

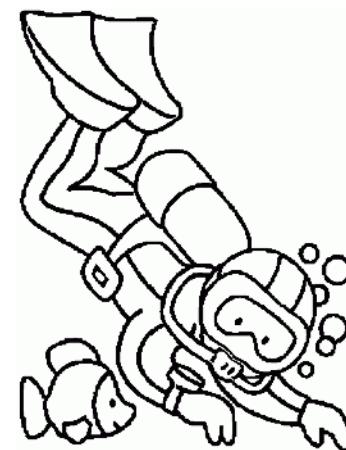
@emma_beauxis

THE “FISH 4 KNOWLEDGE” USE CASE

STUDY 1 - USER NEEDS

STUDY 2 - LAYOUT & INFORMATION COMPLEXITY

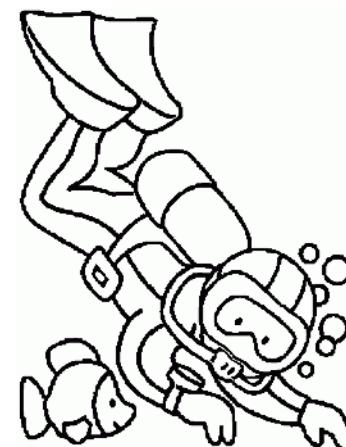
Use Case



We count animals
to study ecosystems

Use Case

Which species
live here?



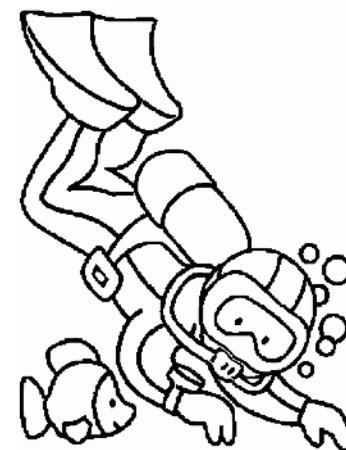
We count animals
to study ecosystems

Use Case

Which species
live here?



How many animals
per species?



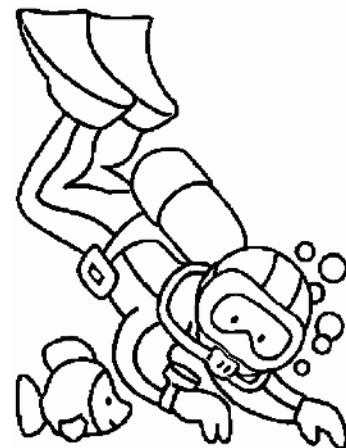
We count animals
to study ecosystems

Use Case



Use computer vision!
You won't disturb us

and you'll save money...

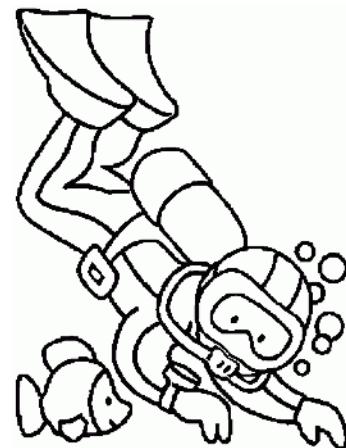


Use Case



Use computer vision!
You won't disturb us

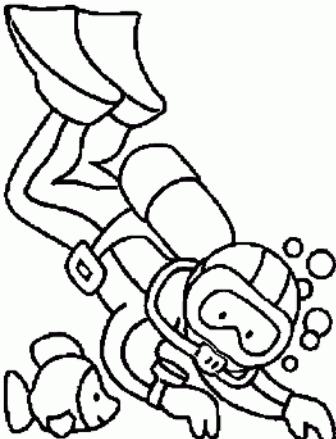
and you'll save money...



...but we need to assess
its scientific validity

Communication Problems

Why should we communicate the uncertainty?



Make informed decisions when choosing and tuning classifiers

Estimate the errors to expect in the classification results



STUDY 1

USER NEEDS

Qualitative Methods: Coding

<https://www.youtube.com/watch?v=DRL4PF2u9XA>

- You can use the steps described in this tutorial to analyze:
 - notes from participatory observations;
 - documents;
 - web pages;
 - or other types of qualitative data.

Suggested reading

Alan Bryman's book: "***Social Research Methods***"
published by Oxford University Press.

Steinar Kvale's and Svend Brinkmann's book
"InterViews: Learning the Craft of Qualitative Research Interviewing" published by SAGE.

The Fish4Knowledge project description:

This project aims at realizing a video analysis tool dedicated to the study of undersea ecosystems. Fixed underwater cameras continuously record videos that are automatically analysed to detect fish species and behaviours.

1. Briefly, what are your scientific research goals and topics of interest?

(if relevant, please name biological patterns, processes or models implied)

2. What information, data or measures do you need to fulfil your goals?

3. How do you collect relevant data (manual methods as well as automated)? What trust or reliability issues do you encounter?

4. What tools do you use to process and analyse those data? What issues do you encounter while using those tools?

5. What would be the 20 most important questions you would ask the Fish4Knowledge tool?

Table 2.1: Questions of the semi-structured interview of marine ecology experts (Section 2.1).

	fish count	species recog.	behavior recog.	body size
1 How many species appear and their abundance and body size in day and night including sunrise and sunset period.	x	x		x
2 How many species appear and their abundance and body size in certain period of time (day, week, month, season or year). Species composition [<i>set of species and relative population sizes</i>] change within one period.	x	x		x
3 Give the rank of above species, i.e., list them according to their abundance or dominance. How many percent are dominant (abundant), common, occasional and rare species.	x	x		
4 Fish colour pattern change and fish behaviour in the night for diurnal fish and in daytime for nocturnal fishes.	x	x	x	
5 Fish activity within one day (24 hours).	x	x	x	
6 Feeding, predator-prey, territorial, reproduction (mating, spawning or nursing) or other social or interaction behavior of various species.	x	x	x	
7 Growth rate of certain species for a certain colony or group of observed fish.	x	x		x
8 Population size change for certain species within a single period of time.	x	x		
9 The relationship of above population size change or species composition change with environmental factors, such as turbidity, current velocity, water temperature, salinity, typhoon, surge or wave, pollution or other human impact or disturbance.	x	x		
10 Immigration or emigration rate of one group of fish inside one monitoring station or one coral head.	x	x		
11 Solitary, pairing or schooling behavior of fishes. [<i>these behavior have different meanings depending on species</i>]	x	x	x	
12 Settle down time or recruitment season [<i>when species stop migrating and start reproducing</i>], body size and abundance for various fish.	x	x		x
13 In certain area or geographical region, how many species could be identified or recognized easily and how many species are difficult. The most important diagnostic characteristics to distinguish some similar or sibling species [<i>species which look-alike</i>].			x	
14 Association [<i>co-occurrence</i>] among different fish species or fish-invertebrates.	x	x		
15 Short term, mid-term or long term fish assemblage [<i>co-occurrence</i>] fluctuation at one monitoring station or comparison between experimental and control stations in MAP. [MPA: <i>Marine Protected Area</i>]	x	x		
16 Comparison of the different study result between using diving observation or underwater real time video monitoring techniques. Or the advantage and disadvantage of using this new technique.	x	x	x	x
17 The difference of using different camera lens and different angle width.	x	x	x	x
18 Is it possible to do the same monitoring in the evening time.	x	x	x	x
19 How to clean the lens and solve the biofouling problem.				
20 Hardware and information technique problem and the possible improvement based on current technology development and how much cost they are.				
21 What is the average body size for species X? How many percent of fish are small, normal or big?	x	x		x
22 What is the number of fish in area X for indicative species related to pollution? [<i>for species which absence is likely due to pollution</i>]	x	x		
23 What is the distribution and number of fish for indicative species of factor X? [<i>for species which presence or absence is likely due to the factor of interest (e.g., water acidity)</i>]	x	x		
24 What is the analysis of factor X impact, using pattern of indicative data Y? [<i>Indicative data include fish counts and behavior observations for indicative species, i.e., species that are known to react to factor X</i>]	x	x	x	
25 What are the areas and periods of time of species X migrations?	x	x		
26 What are the areas and periods of time of species X SPAGS? [<i>SPAGS: Spawning Aggregation Sites, where fish gather to reproduce</i>]	x	x	x	
27 What are the SPAGS periods in area Y?	x	x	x	x

Table 2.4: Information seeking tasks that ecologists would perform with the Fish4Knowledge system. The tasks are reported using participants' own words, in the order they were mentioned, when answering question 5 in Table 2.1. The texts in [...] explain the concepts from the marine ecology domain. The tasks in bold refer to uncertainty or technical issues. The last 4 columns identify high-level information needs (discussed in

	Data Collection Technique	Sampling Method	Uncertainty Issues	Interest in Computer Vision
Case 1	Video Images: baited stereoscopic camera, manual image analysis	Single-point locations	Avoid detecting the same fish multiple times. Few overlaps in fish groups.	To avoid manual image analysis.
Case 2	Video Images: lighted camera held close to deep sea floor, at a constant calibrated distance from seabed, and manual image analysis	Transects (i.e., along a virtual line)	Rare misidentification of species. Cryptic organisms may remain undetected.	To avoid manual image analysis. To reduce expensive use of scientific vessels.
Case 3	Diving Observations with handheld camera for backup purposes	Transects (at varying depths)	Species misidentification. Some species hide from divers. Overlaps in fish groups.	To analyze existing videos. To avoid diving.
Case 4	Experimental Fishery with fish dissection	Single-point locations or transects	Variability of fish catch albeit identical experimental conditions.	Excluded, due to unsupported measurements and uncertainty issues
Case 5	Commercial Fishery: data from the North-Sea fish market	Dependent on commercial fisheries	Variability of fish catch. Targets only commercial species. Misidentifies uncommon species.	To compensate the biases in the market-dependent sampling conditions
Case 6	Diving Observations	Single-point locations or transects	Species misidentification. Some species are hiding from divers. Overlaps in fish groups.	To avoid diving.
Case 7	Video Images & Commercial Fishery: onboard video monitoring of fish discarded during fish processing	Dependent on equipment available onboard	Misidentification of species and non-fish objects.	Experimented in 2013, needs improvement.

Table 2.2: Summary of 7 typical use cases of fish population monitoring for ecology research.

	Research Topic				Information Need			
	Population Dynamics	Migration	Reproduction	Trophic Systems	Fish Count	Species Recognition	Body Size	Other
Case 1	x	x			x	x	x	
Case 2	x			x	x	x	x	Other organisms
Case 3	x			x	x	x	x	
Case 4	x	x	x	x	x	x	x	Weight, Bone size, Stomach content, Chemicals
Case 5	x	x	x		x	x	x	Weight
Case 6	x		x	x	x	x	x	Behavior
Case 7	x				x	x		

Table 2.3: High-level information needs drawn from the use cases in Section 2.2 and Table 2.2.

	Fish Count	Species Recognition	Behavior Recognition	Body Size
Research Topic				
Population Dynamics	mandatory	mandatory	optional	important
Migration	mandatory	mandatory	optional	optional
Reproduction	mandatory	mandatory	important	important
Trophic Systems	mandatory	mandatory	important	important
Data Collection Technique				
Experimental Fishery (Case 4)	+	+/++ ¹	-	+
Commercial Fishery (Cases 5, 7)	+	+	-	+
Diving Observation (Cases 3, 6)	+	+	++	+
Manual Image Analysis (Cases 1, 2, 3, 7)	+	+	+	-/+ ²
Computer Vision	+	+	-/+ ³	-/+ ²

The signs indicate whether data collection techniques: - cannot supply the information, + can supply the information, ++ can supply the most precise information.

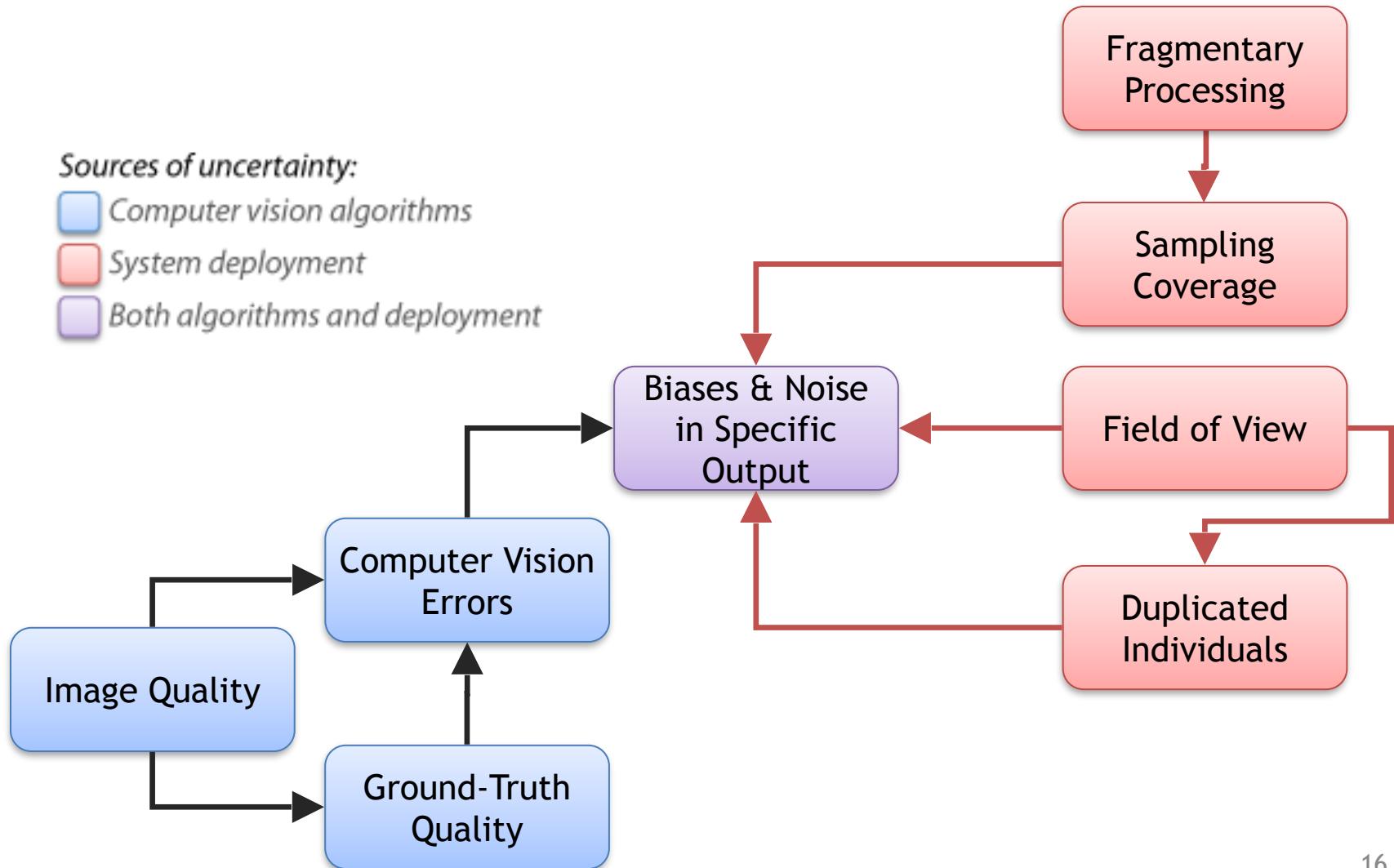
¹ Fish dissection, sometimes performed after experimental fishing, is the most accurate technique for differentiating fish species that are visually similar.

² Information supplied if stereoscopic vision or calibrated distance camera-background are available.

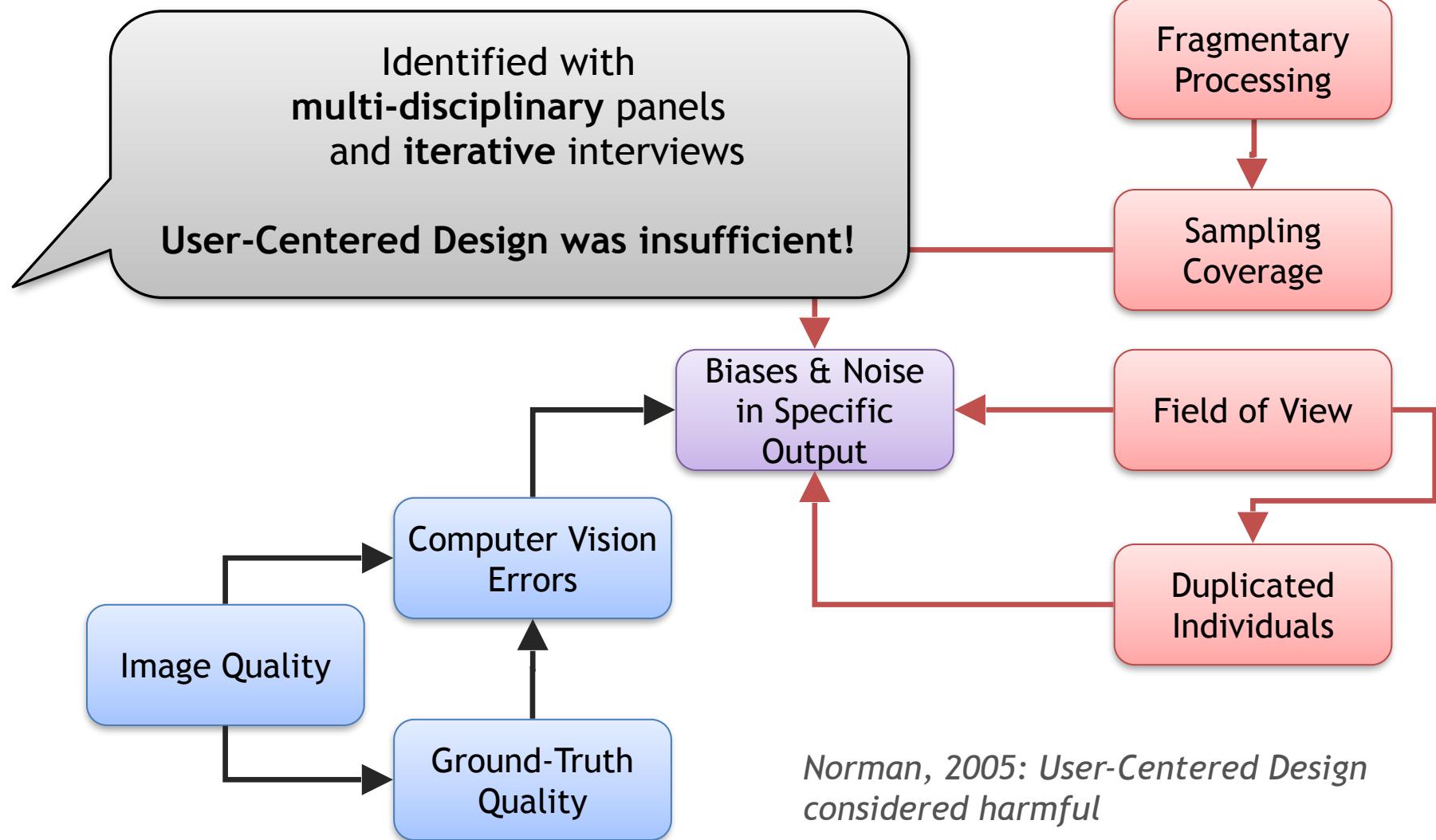
³ The state-of-the-art does not fully address the wide scope of fish behavior variety.

Table 2.5: Information required for the main topics of study, and ability of data collection techniques to provide this information.

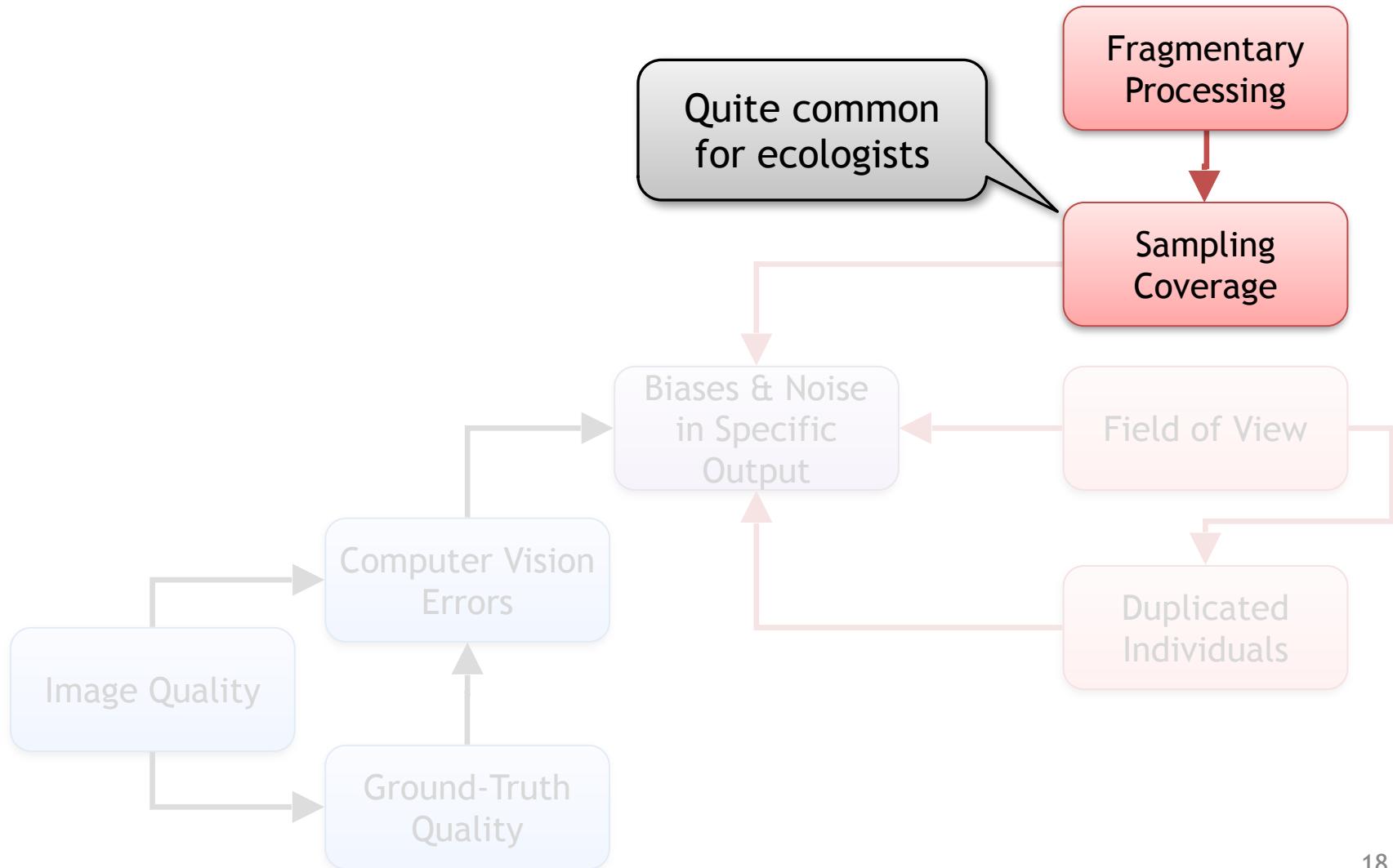
Interactions of Uncertainty Factors



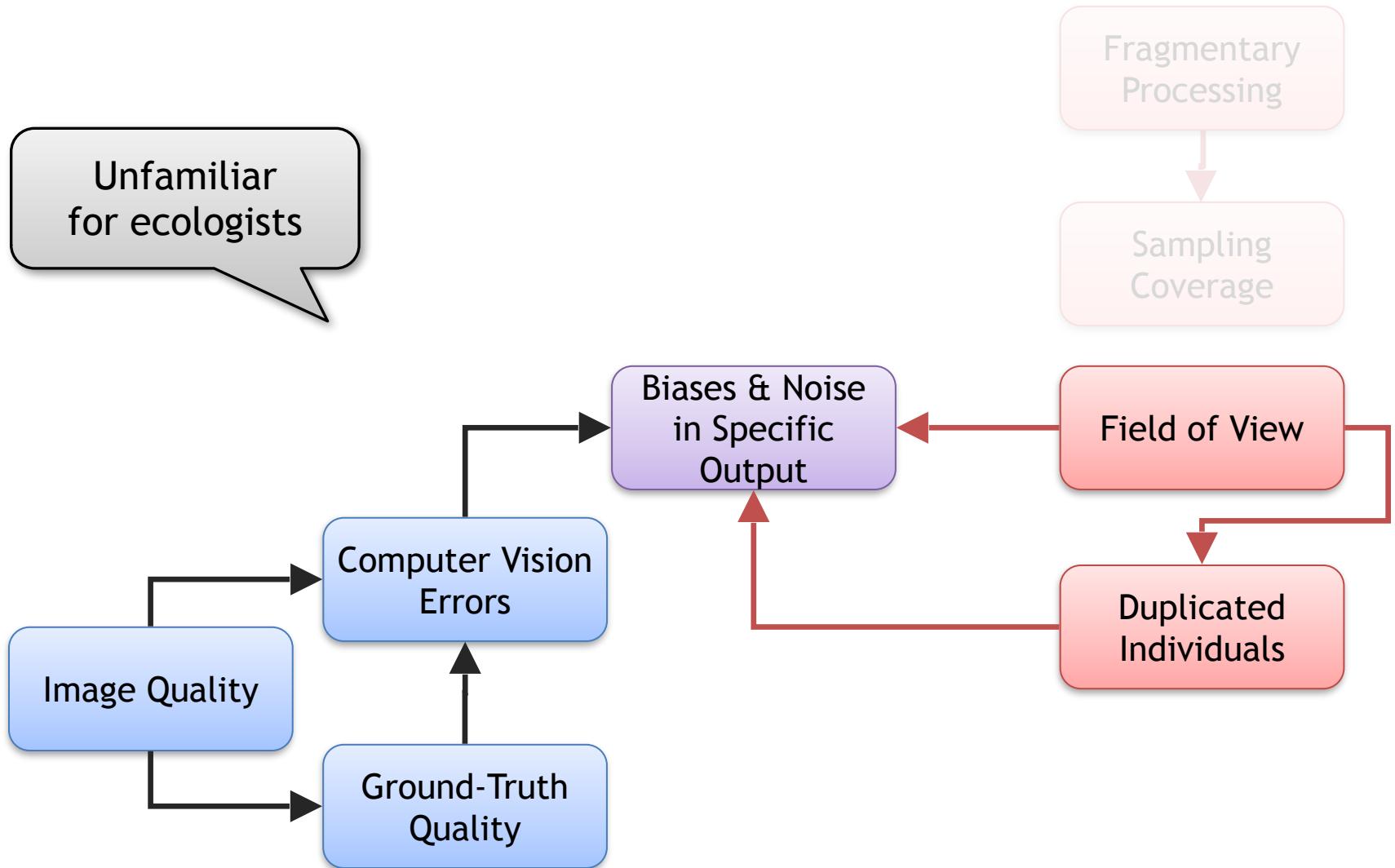
Interactions of Uncertainty Factors



Interactions of Uncertainty Factors



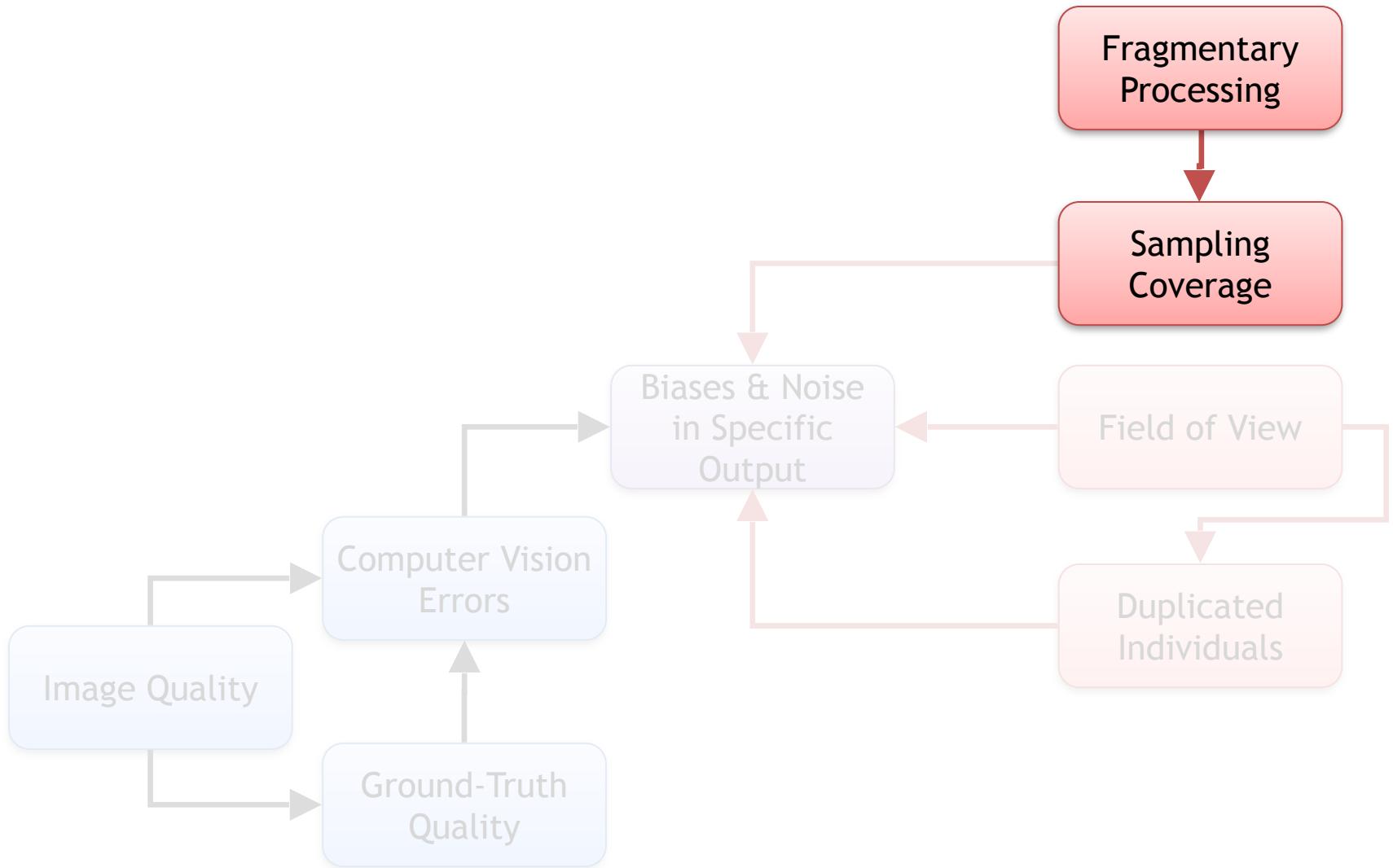
Interactions of Uncertainty Factors





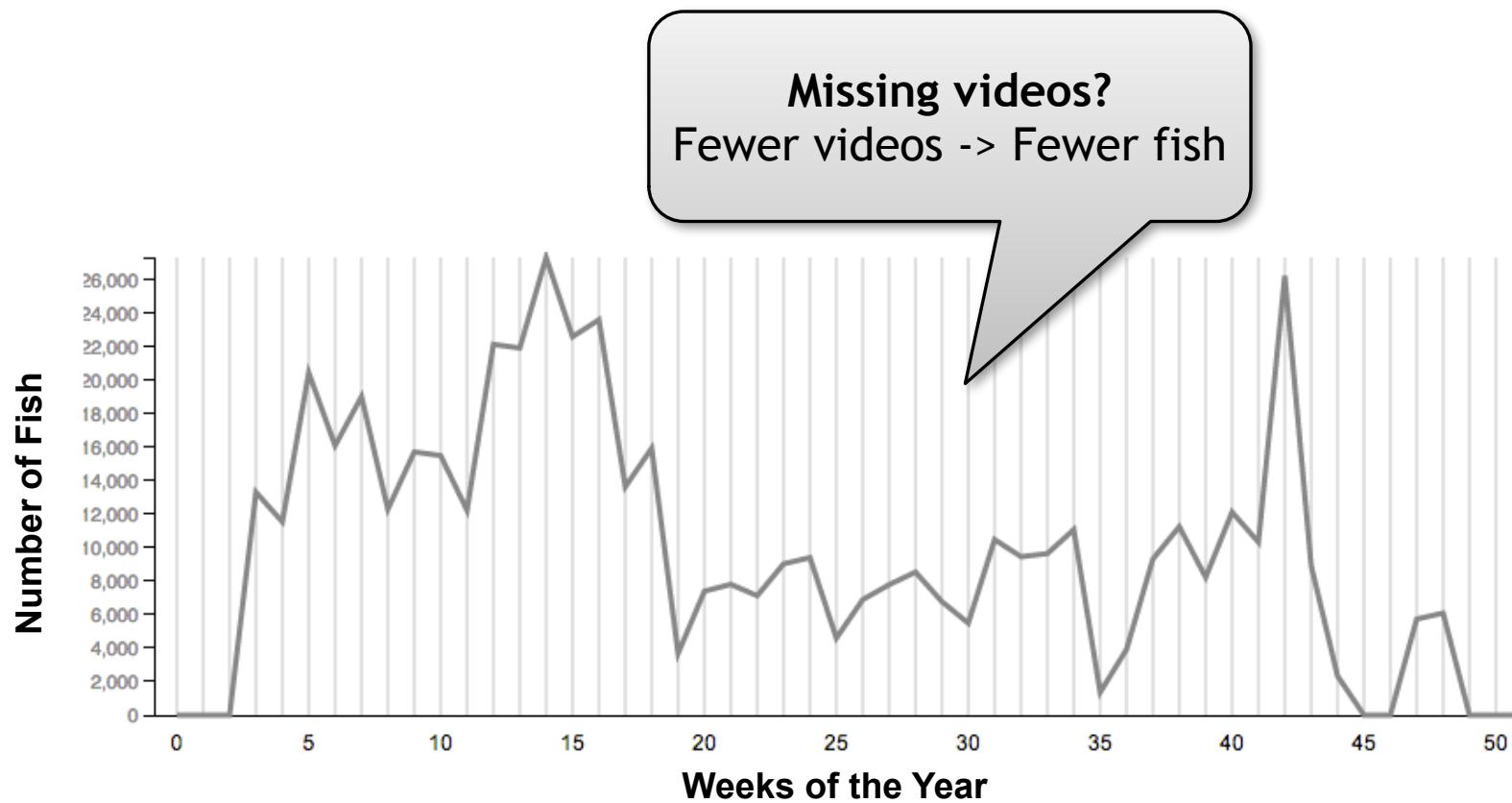
STUDY 2

LAYOUT & INFO COMPLEXITY



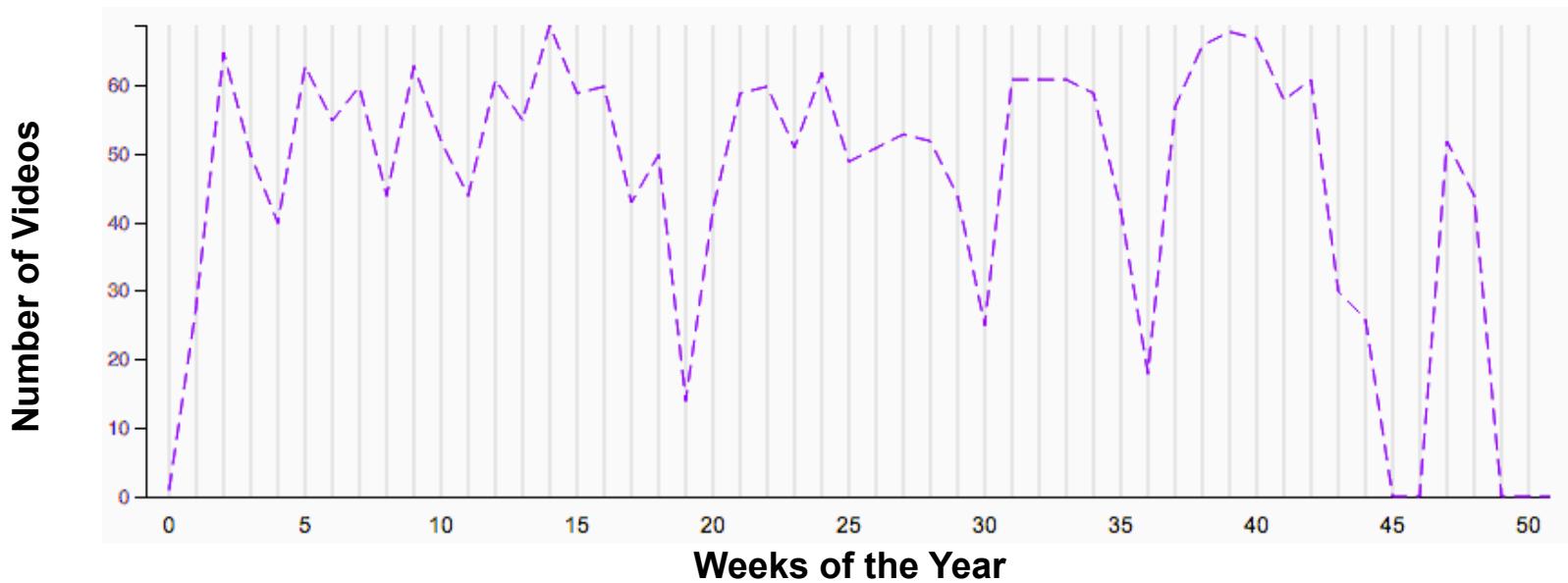
[1] Beauxis-Aussalet, 2015. Supporting Non-Experts' Awareness of Uncertainty: Negative Effects of Simple Visualizations in Multiple Views. European Conference on Cognitive Ergonomics. <https://ir.cwi.nl/pub/23433>

Fragmentary Processing



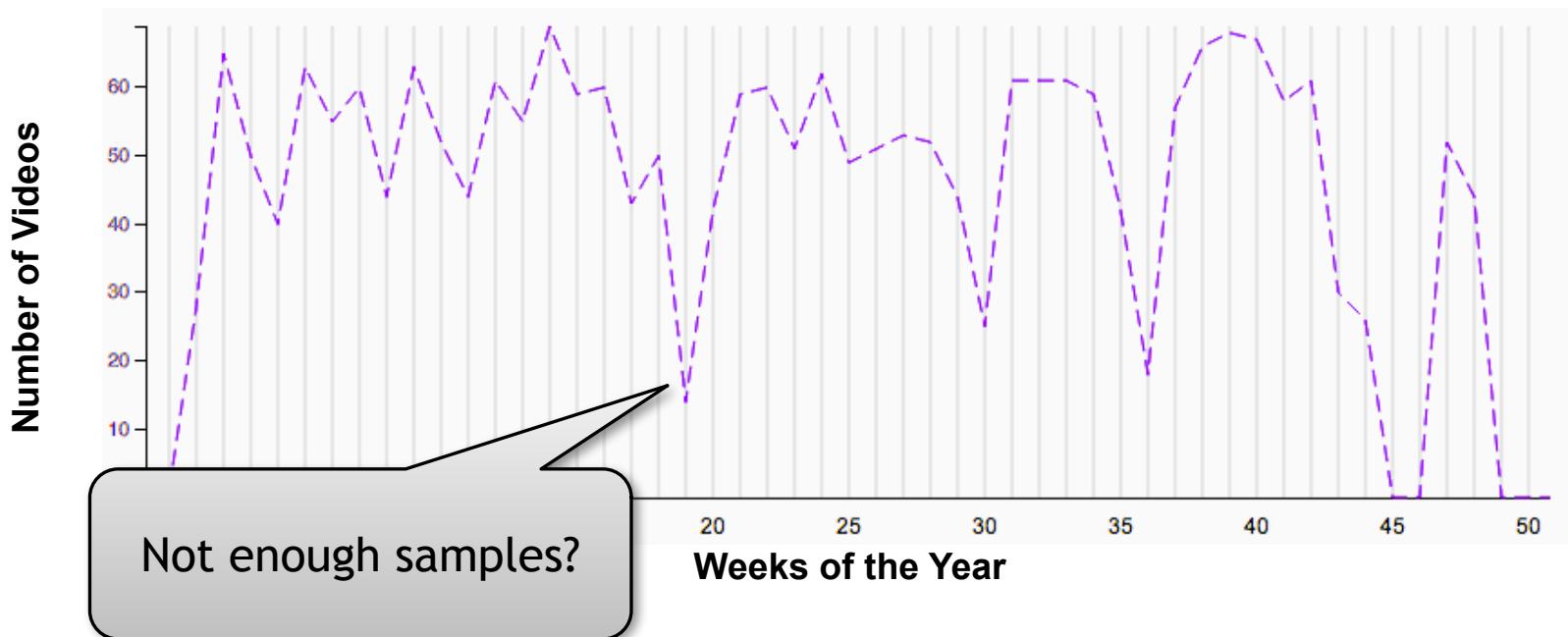
Fragmentary Processing

We investigated the use of **Number of Video Samples** to measure this uncertainty



Fragmentary Processing

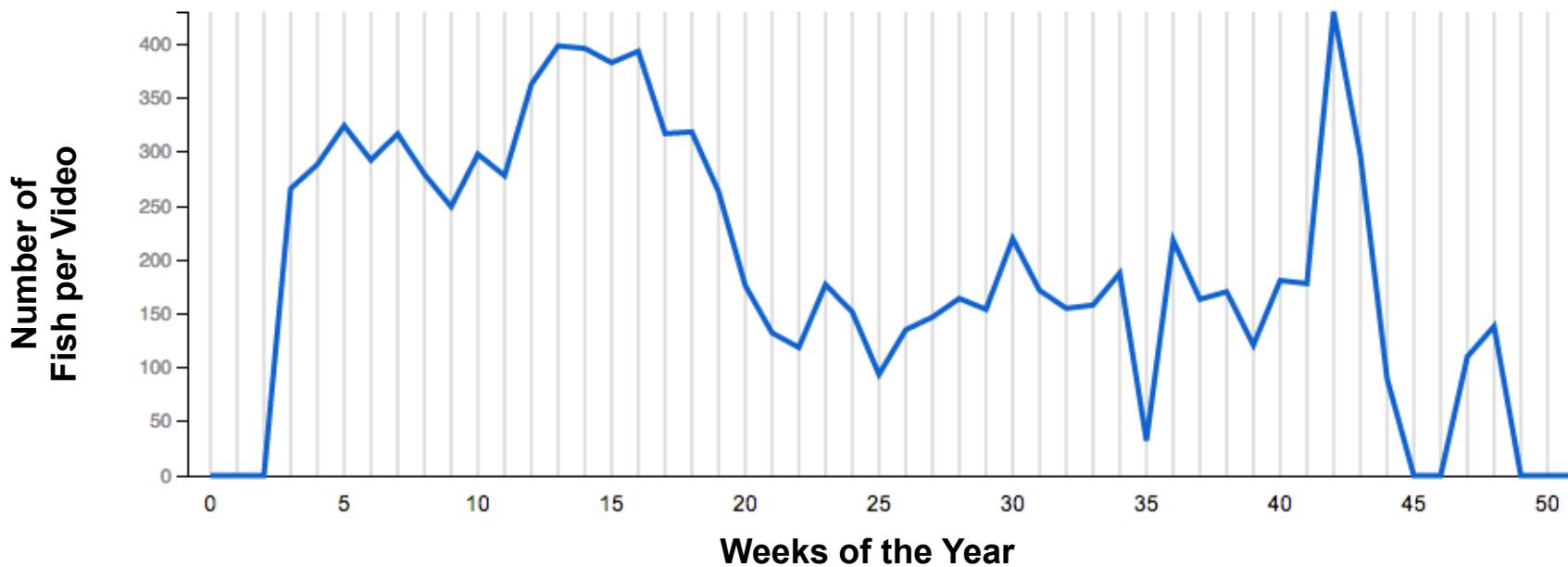
We investigated the use of **Number of Video Samples** to measure this uncertainty



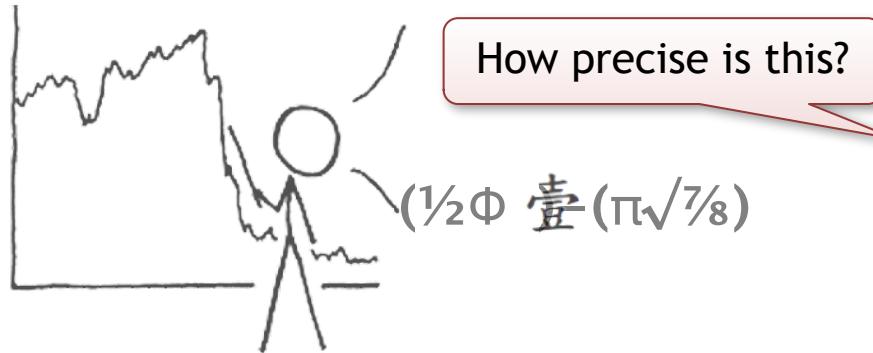
Fragmentary Processing

We investigated the use of **Number of Video Samples** to measure this uncertainty

...together with **Average Number of Fish per Video**

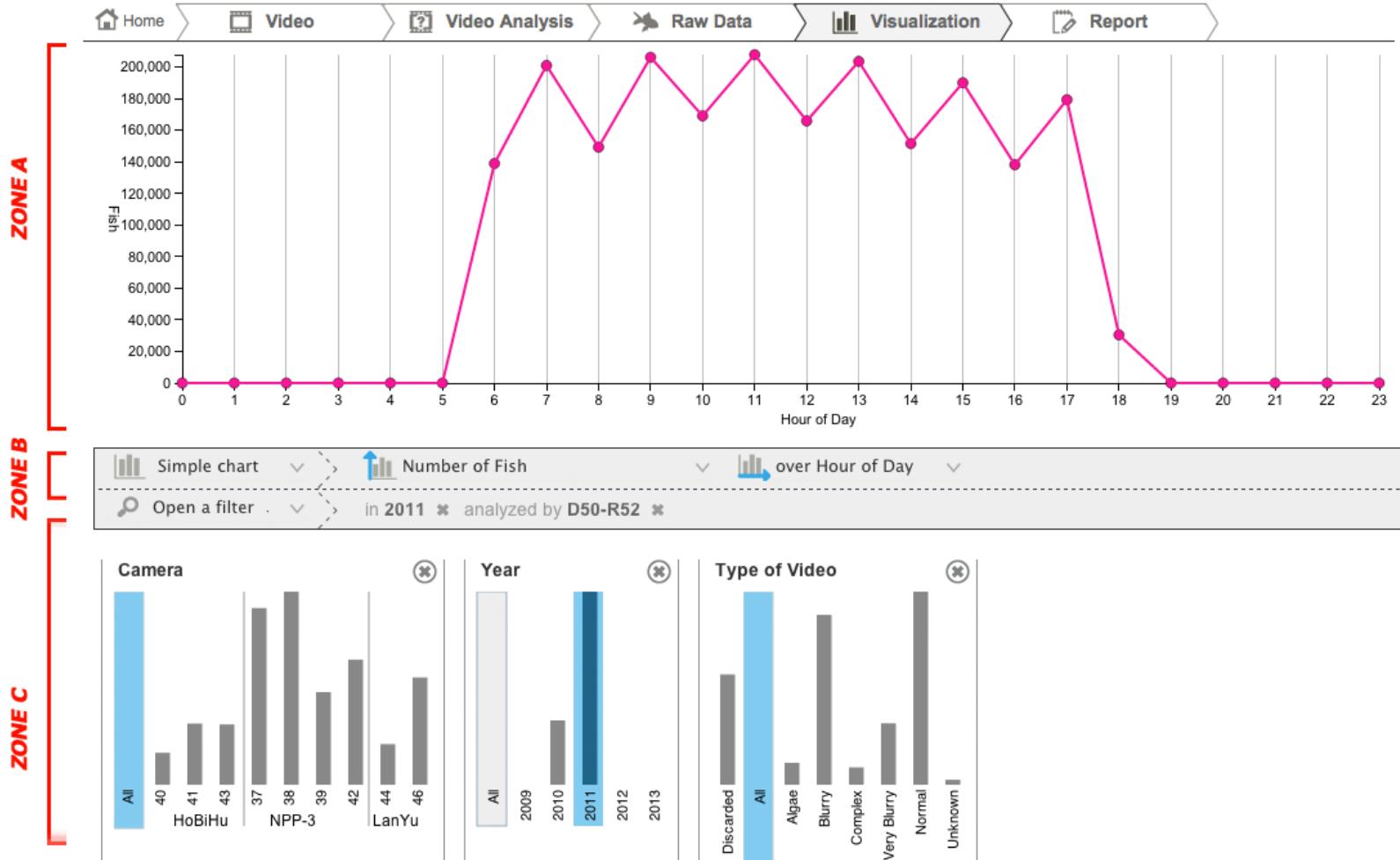


Here the Octopus appeared.

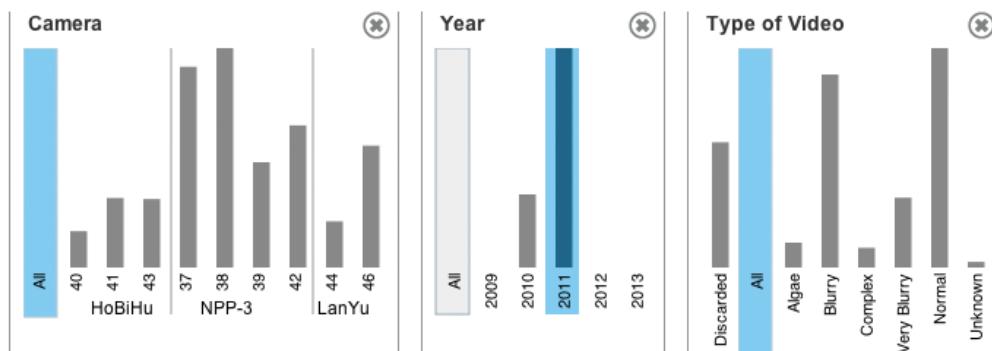
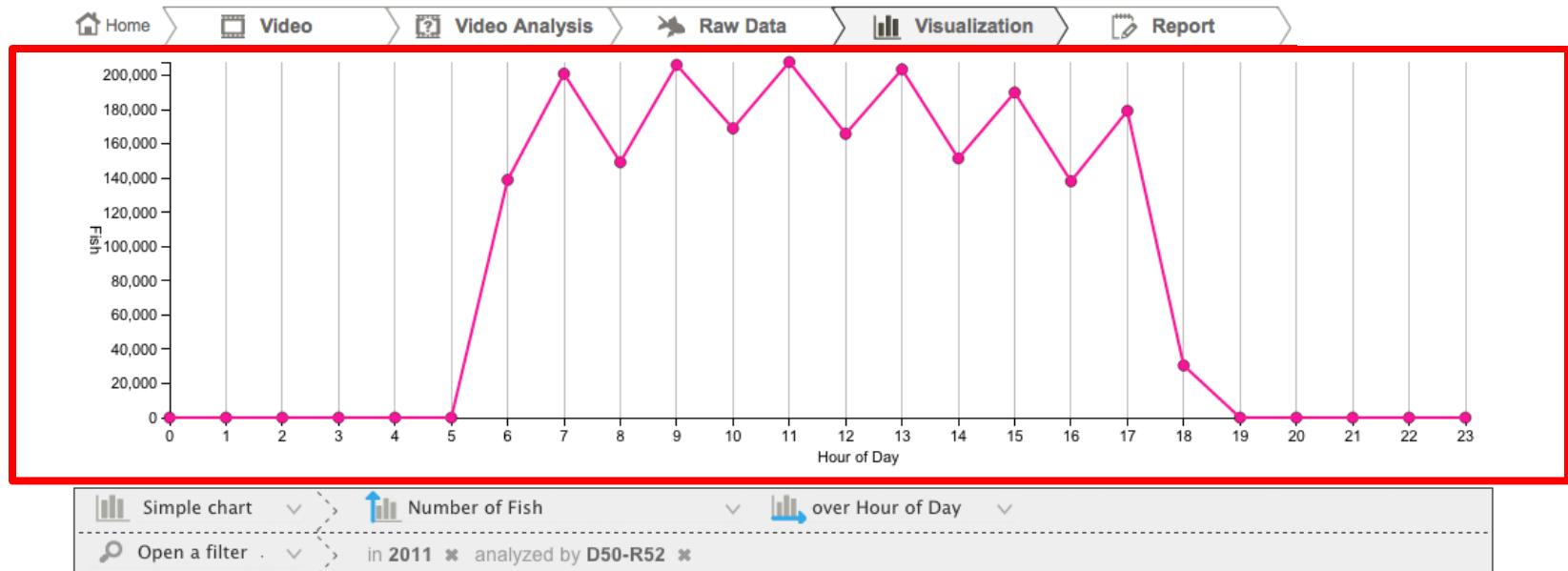


How to visualize the complex information?

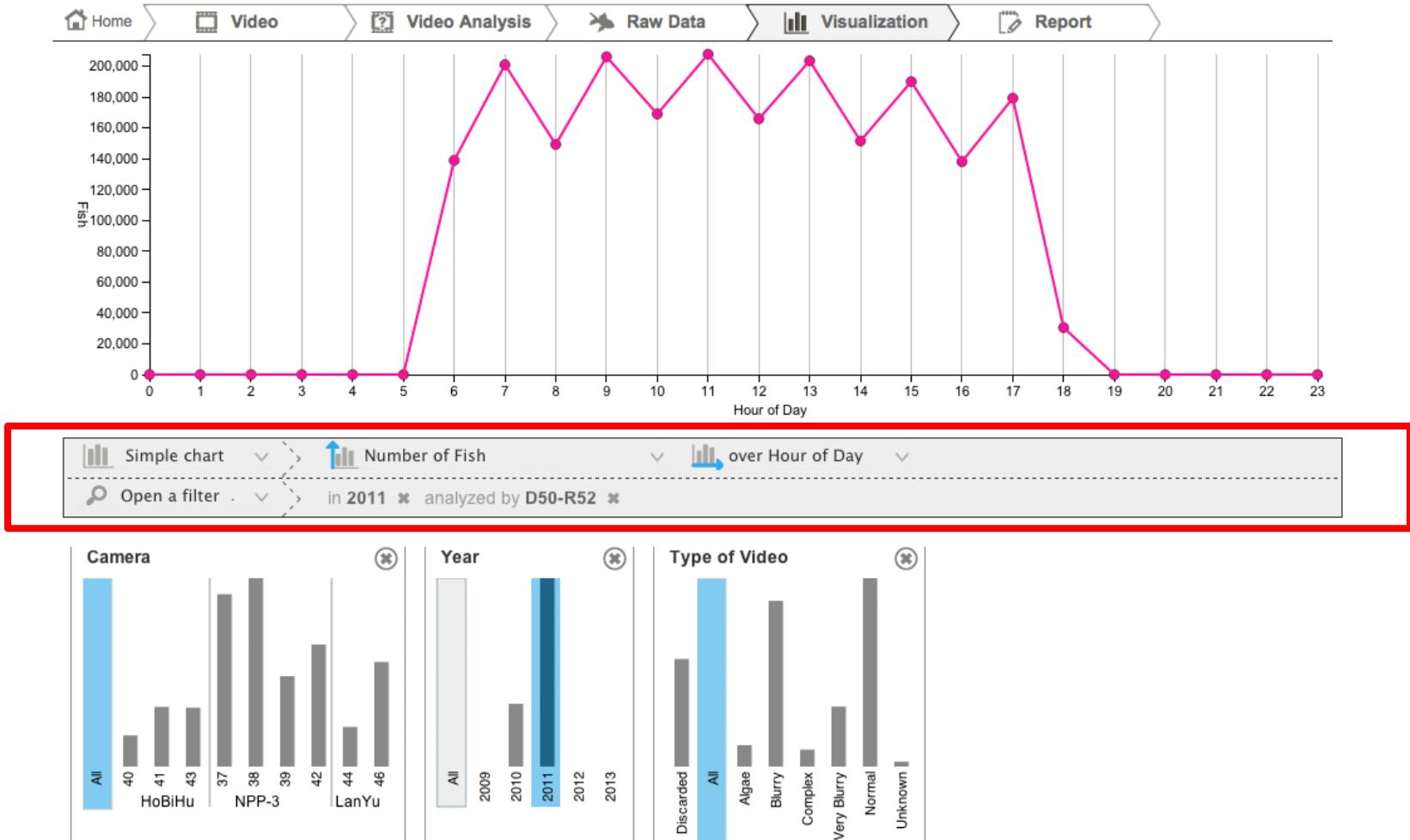
Simple Graphs in Multiple Views



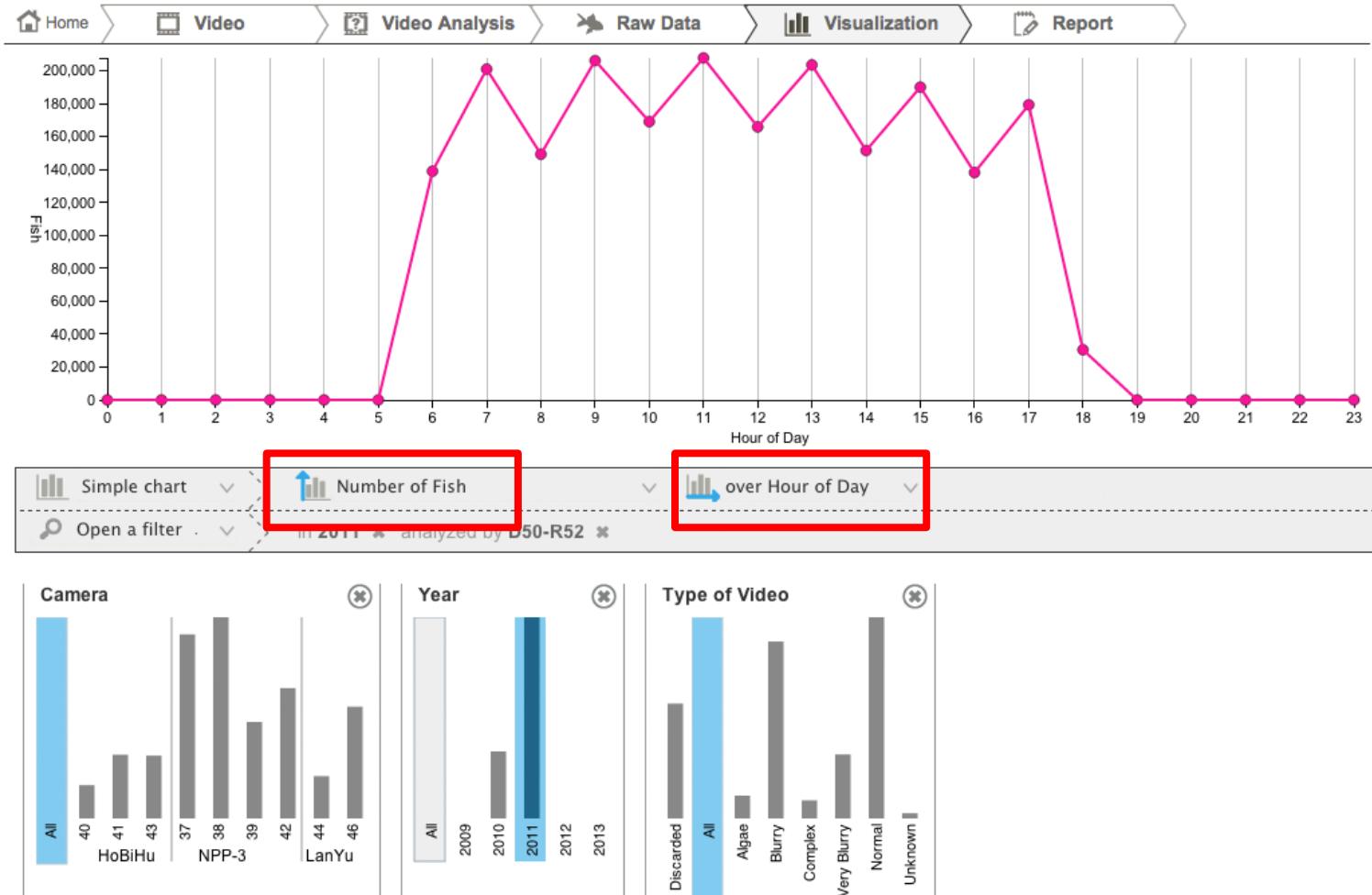
Main Graph



Controls over Main Graph



Swap the Graph Axes



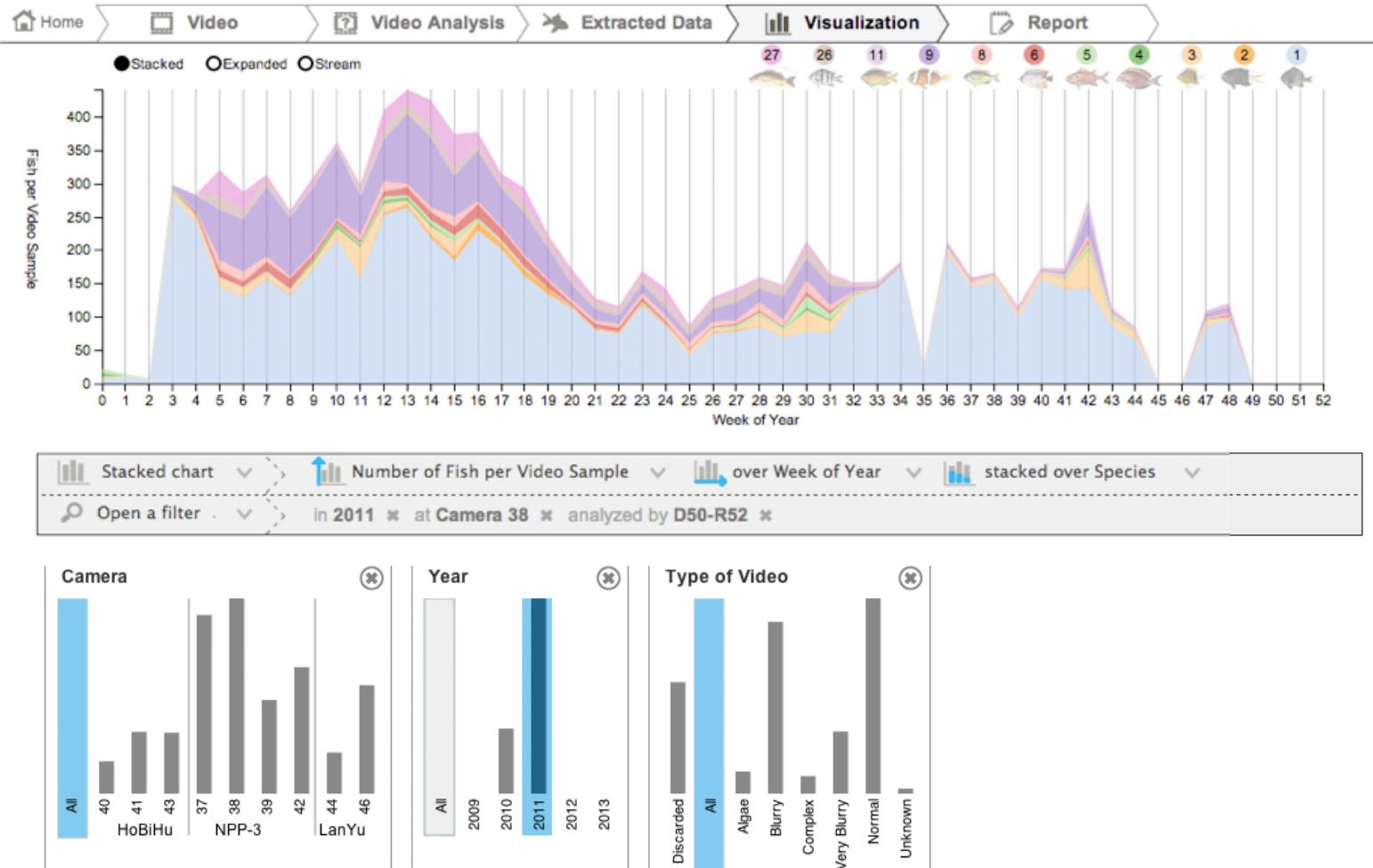
Change the Type of Graph...



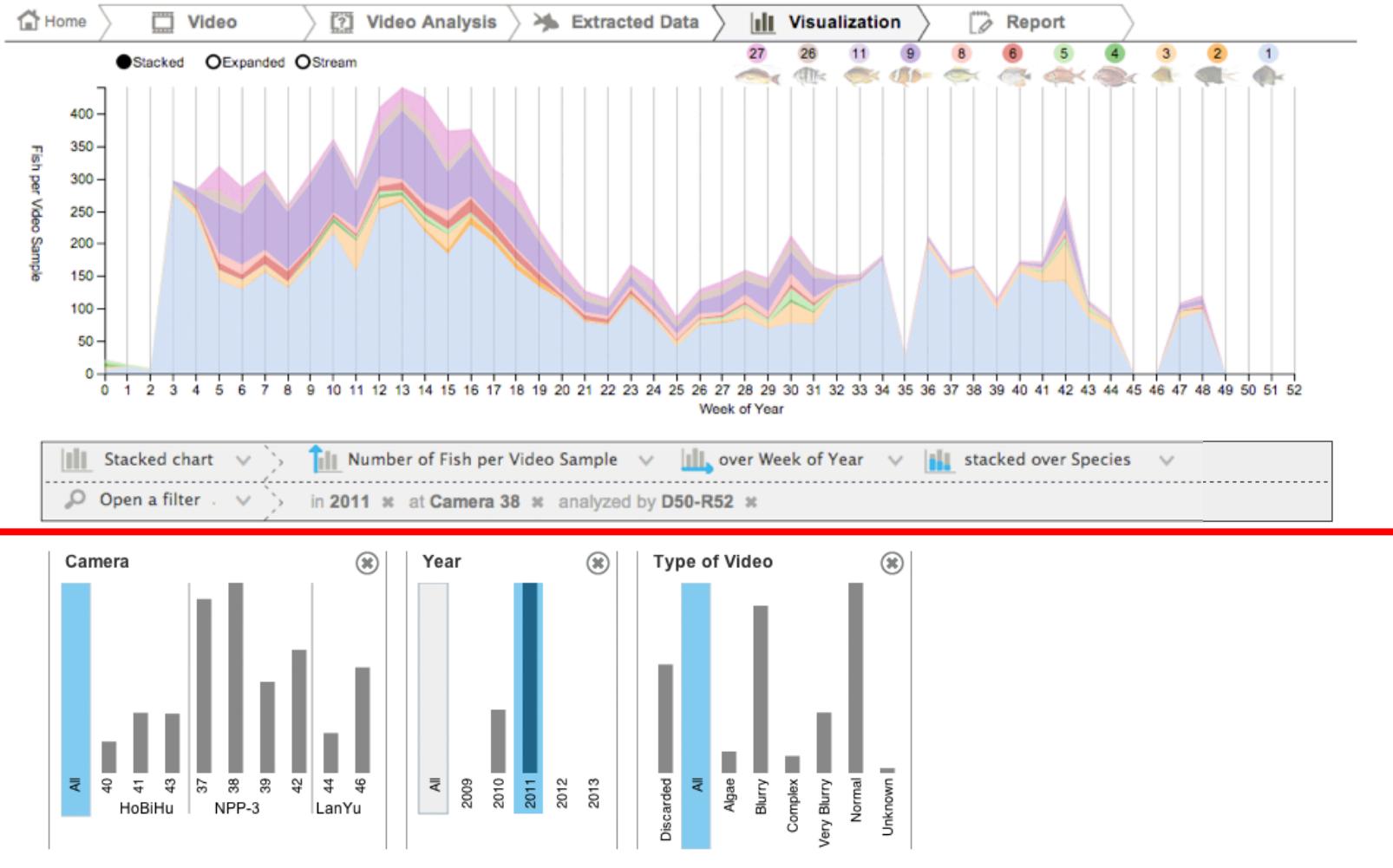
...adding a 3rd Dimension to the Graph



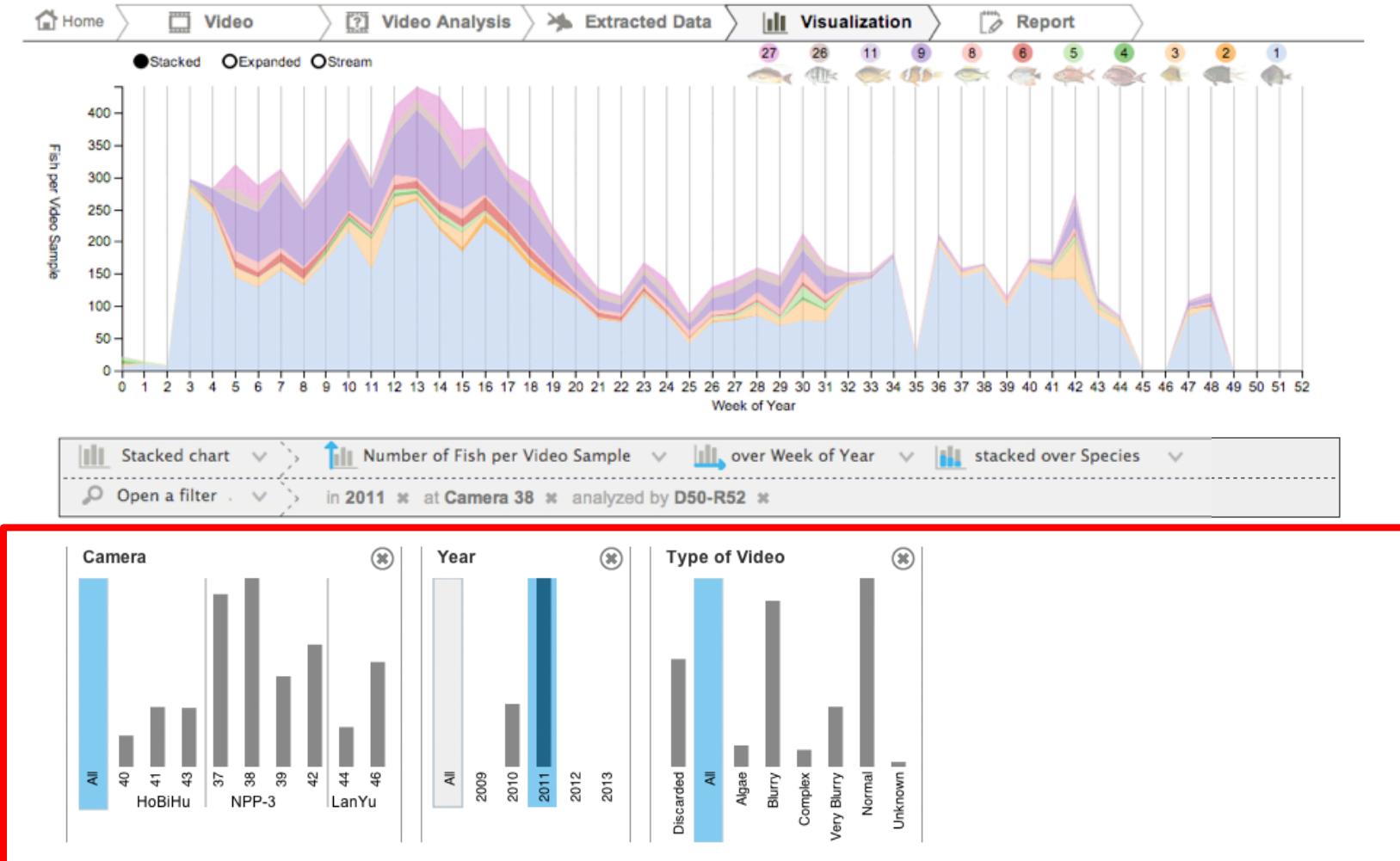
Add a 3rd Dimension to the Main Graph



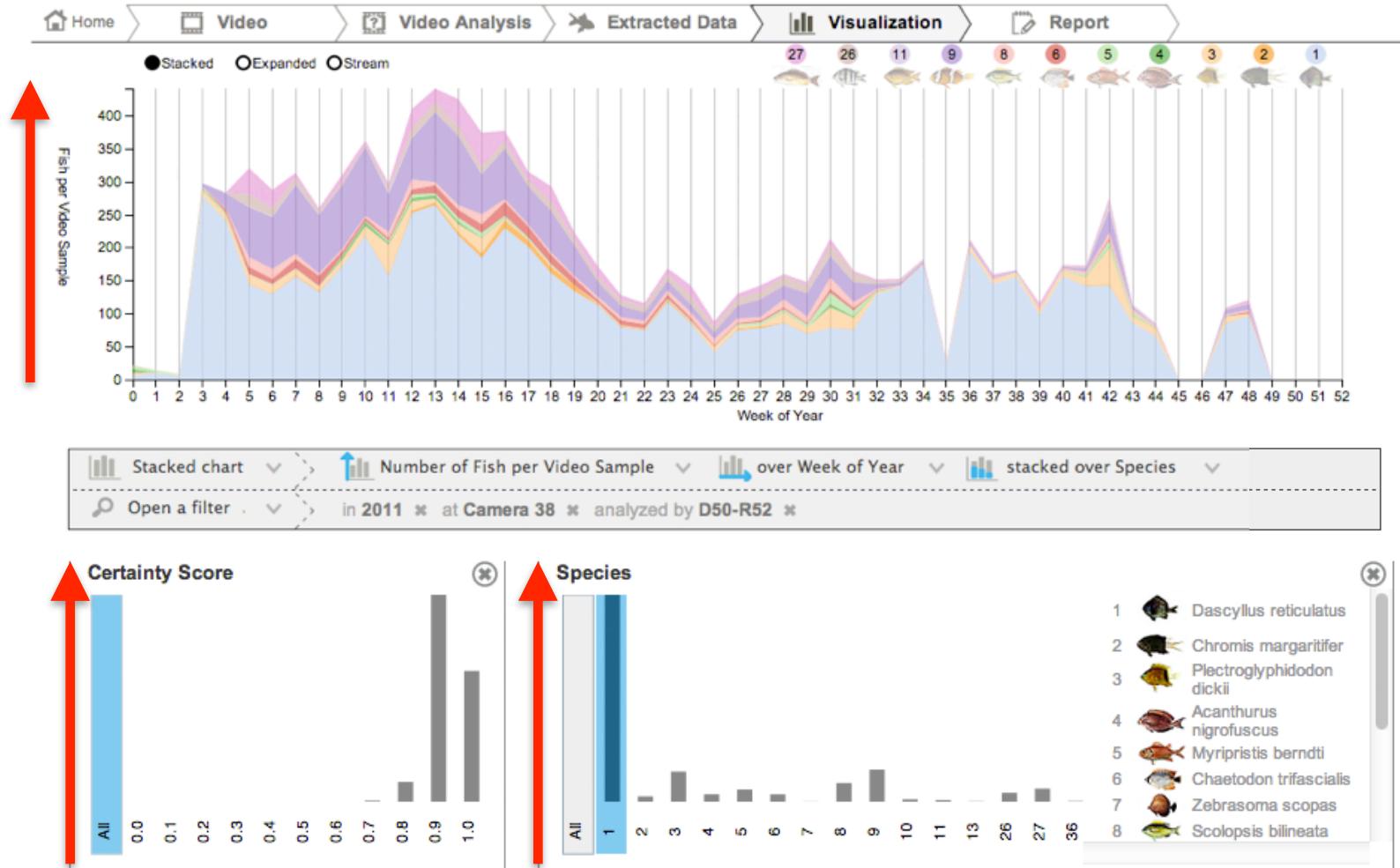
Select the Dataset of Interest...



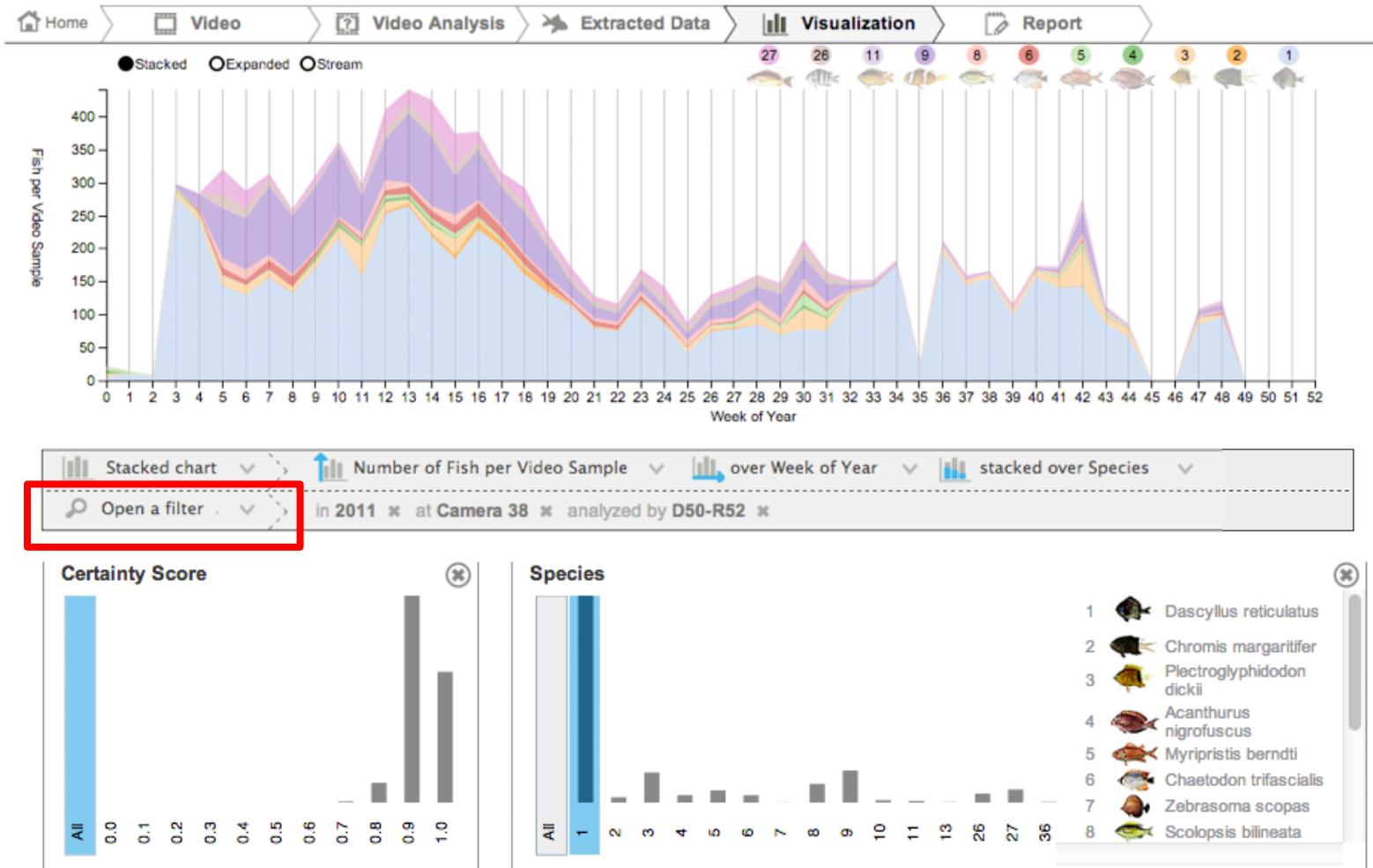
...and Overview the Data



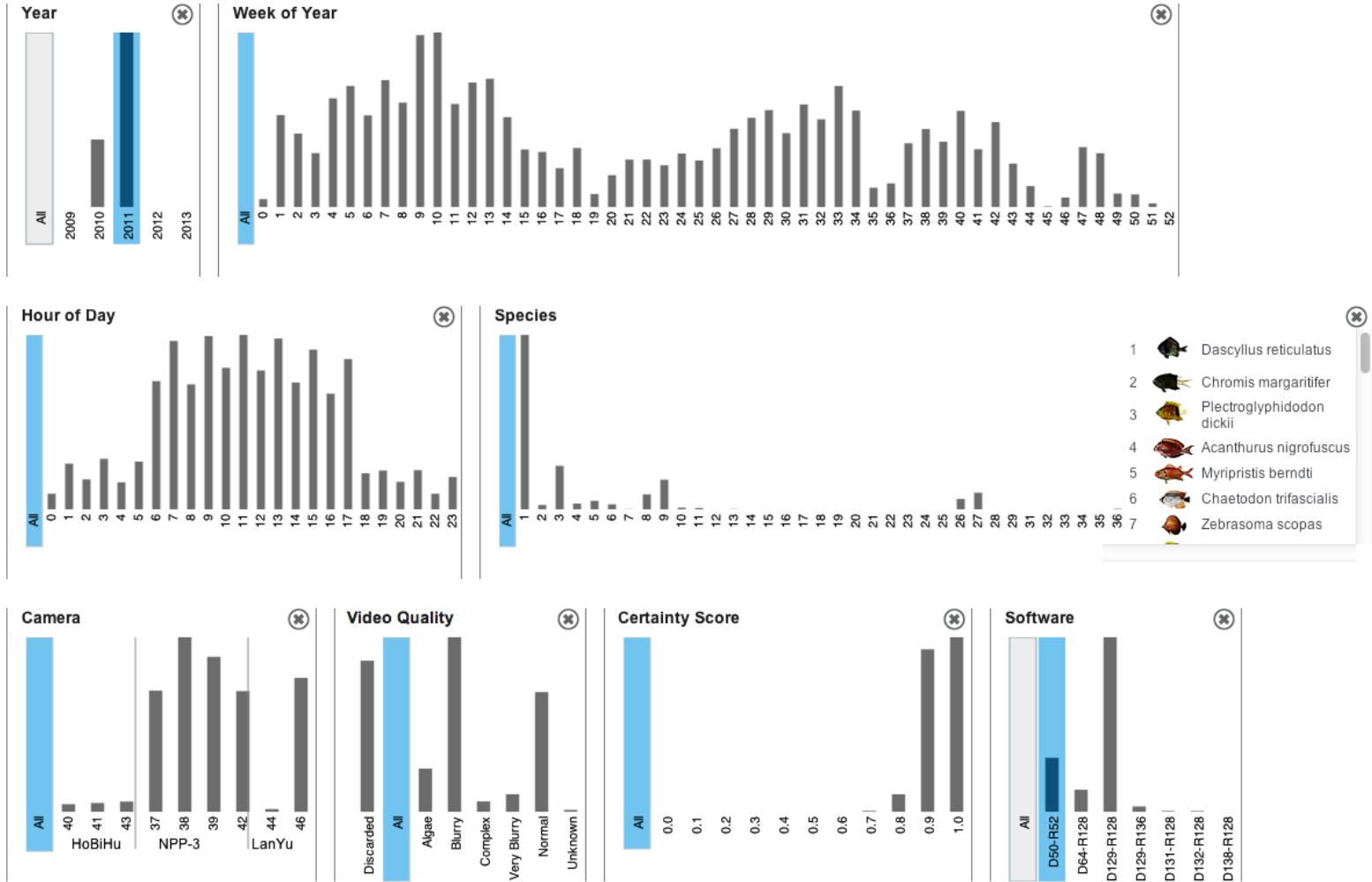
Use the Same Y-Axis



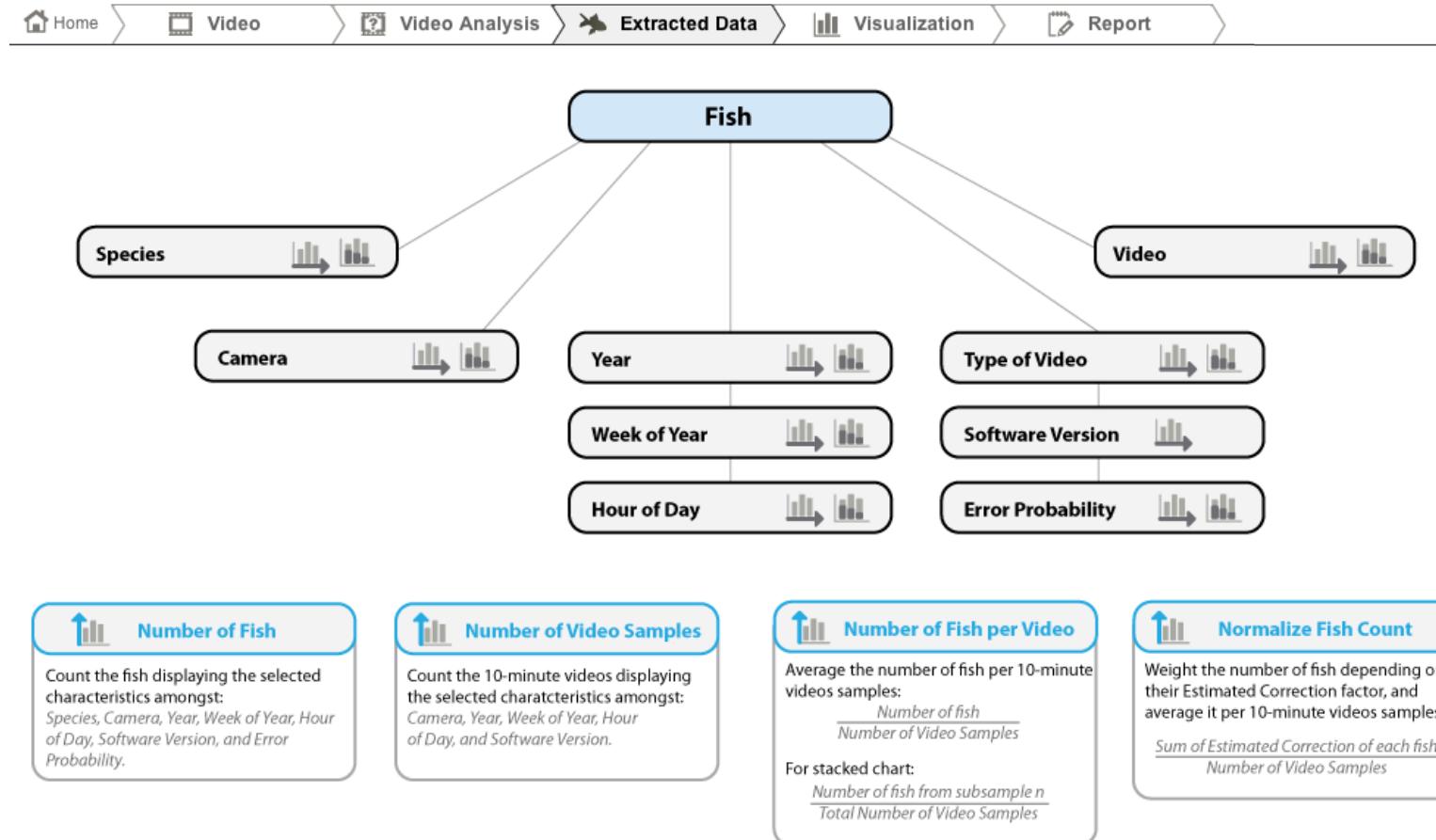
Select Data Dimensions to Display



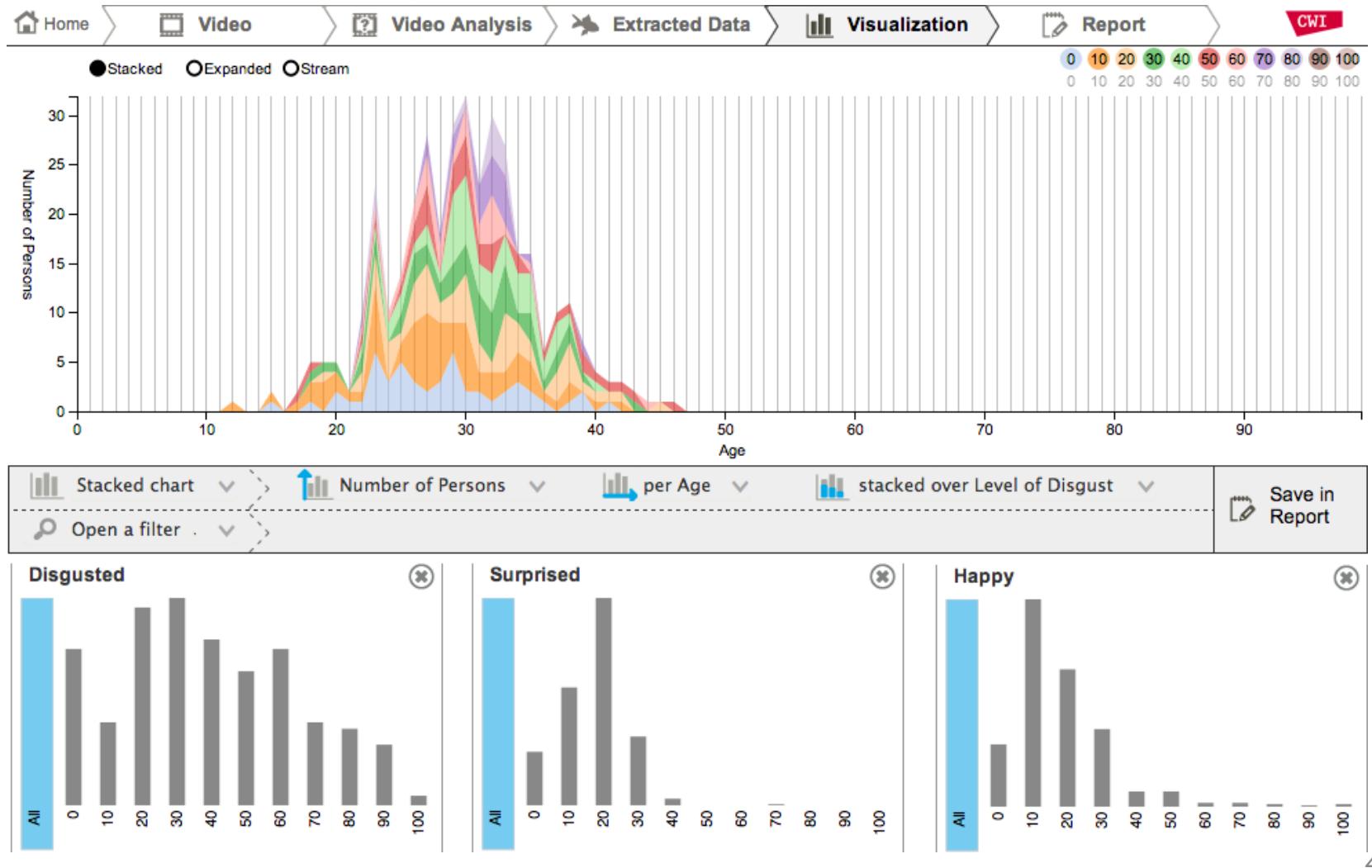
Handle Multiple Dimensions



Handle Multiple Dimensions



Other Applications



Other Applications

Home > Video > Video Analysis > Extracted Data > Visualization > Report

Videos

Camera37 (387 videos)

2011-02-20 07:00 (140 fish)
2011-02-20 08:00 (127 fish)
2011-02-20 09:00 (101 fish)
2011-02-20 10:00 (117 fish)
2011-02-20 11:00 (147 fish)

Camera38 (423 videos)

2011-02-20 07:00 (243 fish)
2011-02-20 10:00 (117 fish)
2011-02-20 11:00 (147 fish)

Show more

2011-01-01 17:00
 2011-01-01 17:00:00
2:05

Species counts in current video

Dascyllus reticulatus (6)
Plectroglyphidodon dickii (4)
Acanthurus nigrofasciatus (6)
Myripristis berndti (5)
Scolopsis bilineata (1)

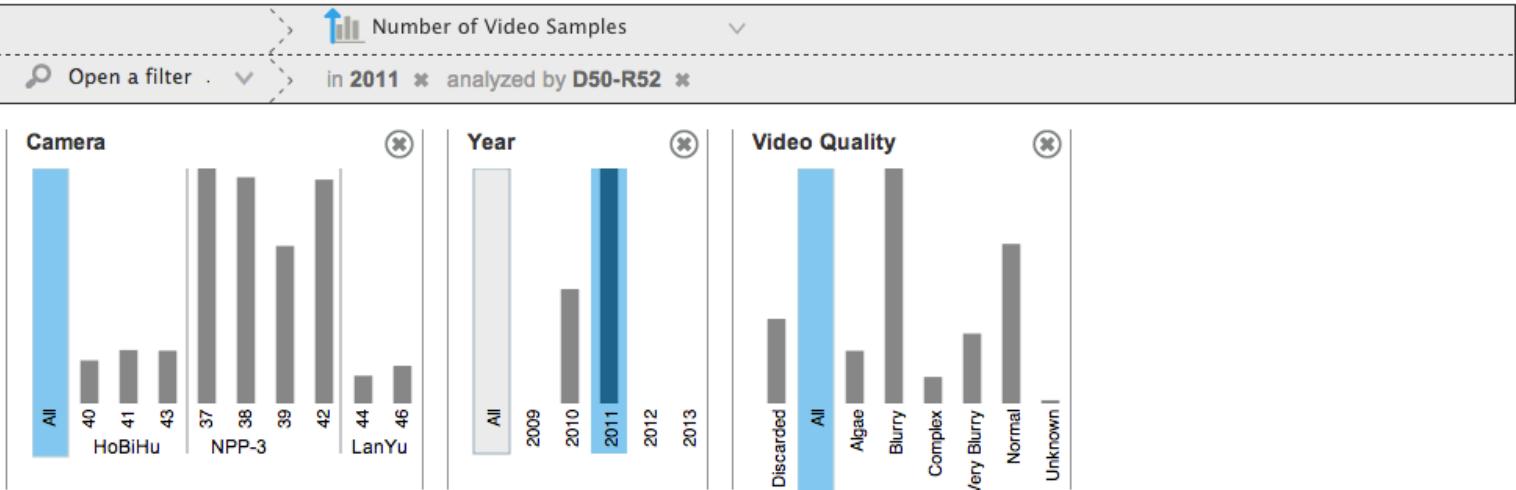
Number of Video Samples

Open a filter in 2011 analyzed by D50-R52

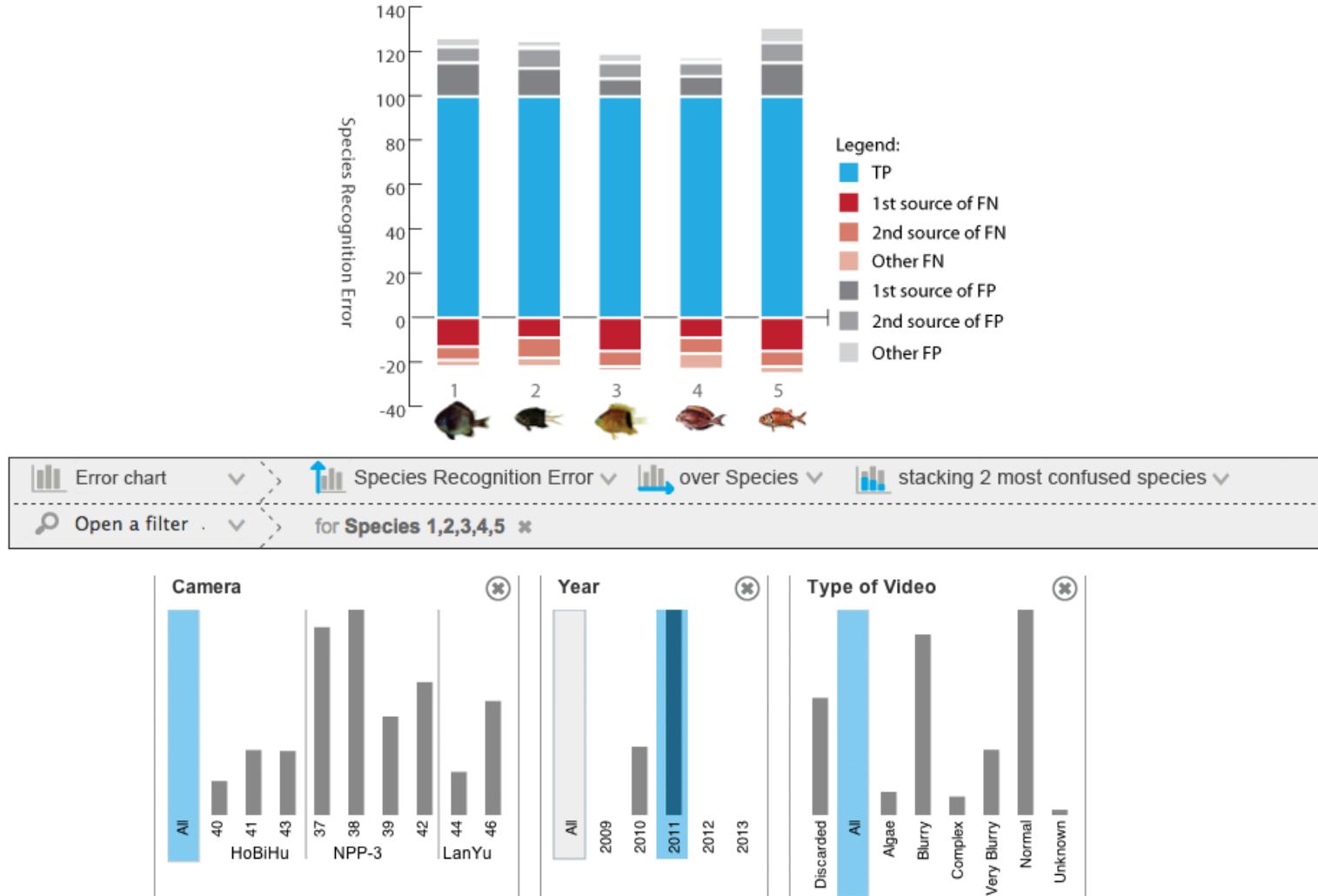
Camera: All, 40, 41, 43, 37, 38, 39, 42, 44, 46; HoBiHu, NPP-3, LanYu

Year: All, 2009, 2010, 2011, 2012, 2013

Video Quality: Discarded, All, Algae, Blurry, Complex, Very Blurry, Normal, Unknown



Other Applications





How to evaluate the interface?

What did we investigate?

Does our design support information seeking tasks?

What did we investigate?

Does our design support information seeking tasks?

Did users find
the right answer?

What did we investigate?

Does our design support information seeking tasks?

Did users find
the right answer?

With the correct
level of confidence?

What did we investigate?

Does our design support information seeking tasks?

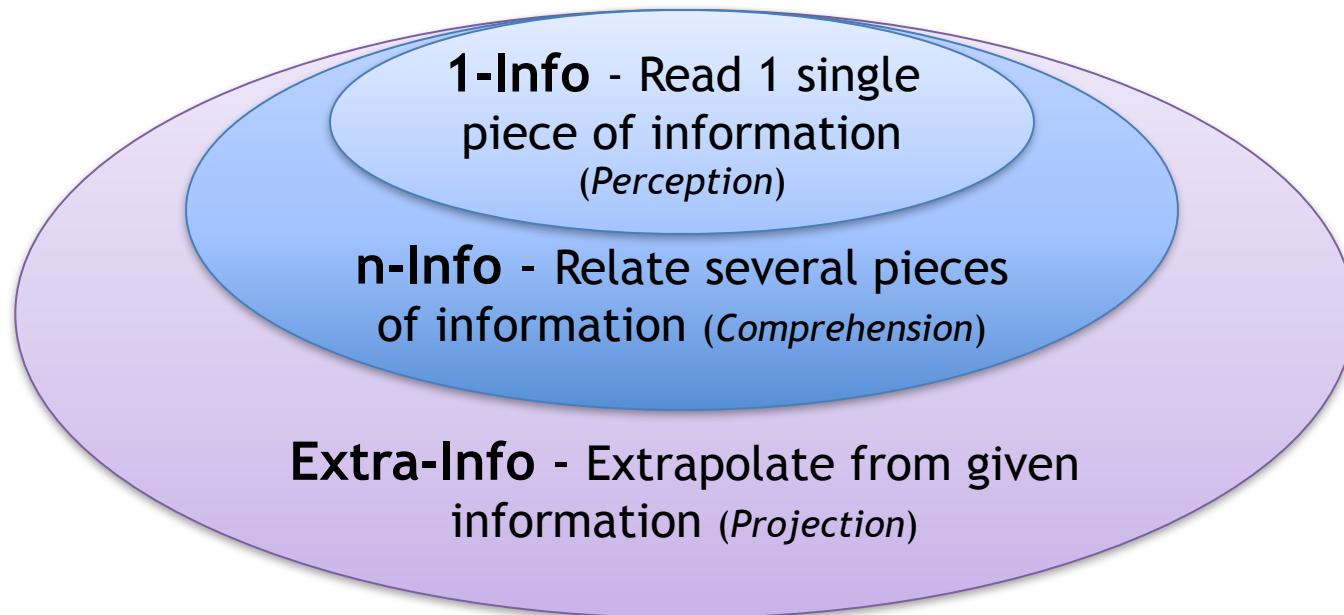
Did users find
the right answer?

With the correct
level of confidence?

Did users encounter
usability issues?

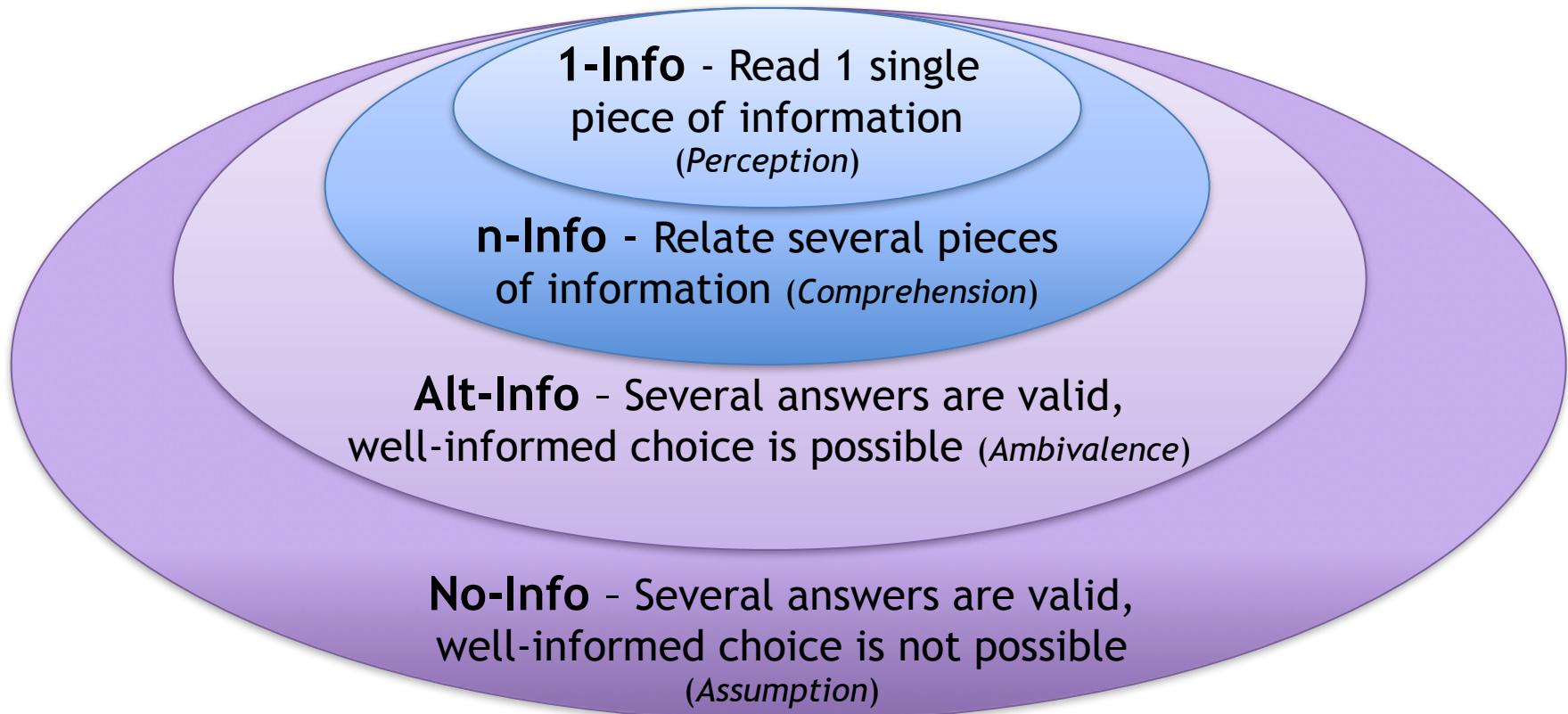
Differentiate Levels of Complexity

Does our design support simple to complex tasks?



Differentiate Levels of Complexity

...including the complexity of **uncertainty**?



Differentiate Levels of Complexity

“What is the number of fish for the week 12?”

“For which camera are we counting the fish?”

1-Info - Read 1 single piece of information
(*Perception*)

n-Info - Relate several pieces of information
(*Comprehension*)

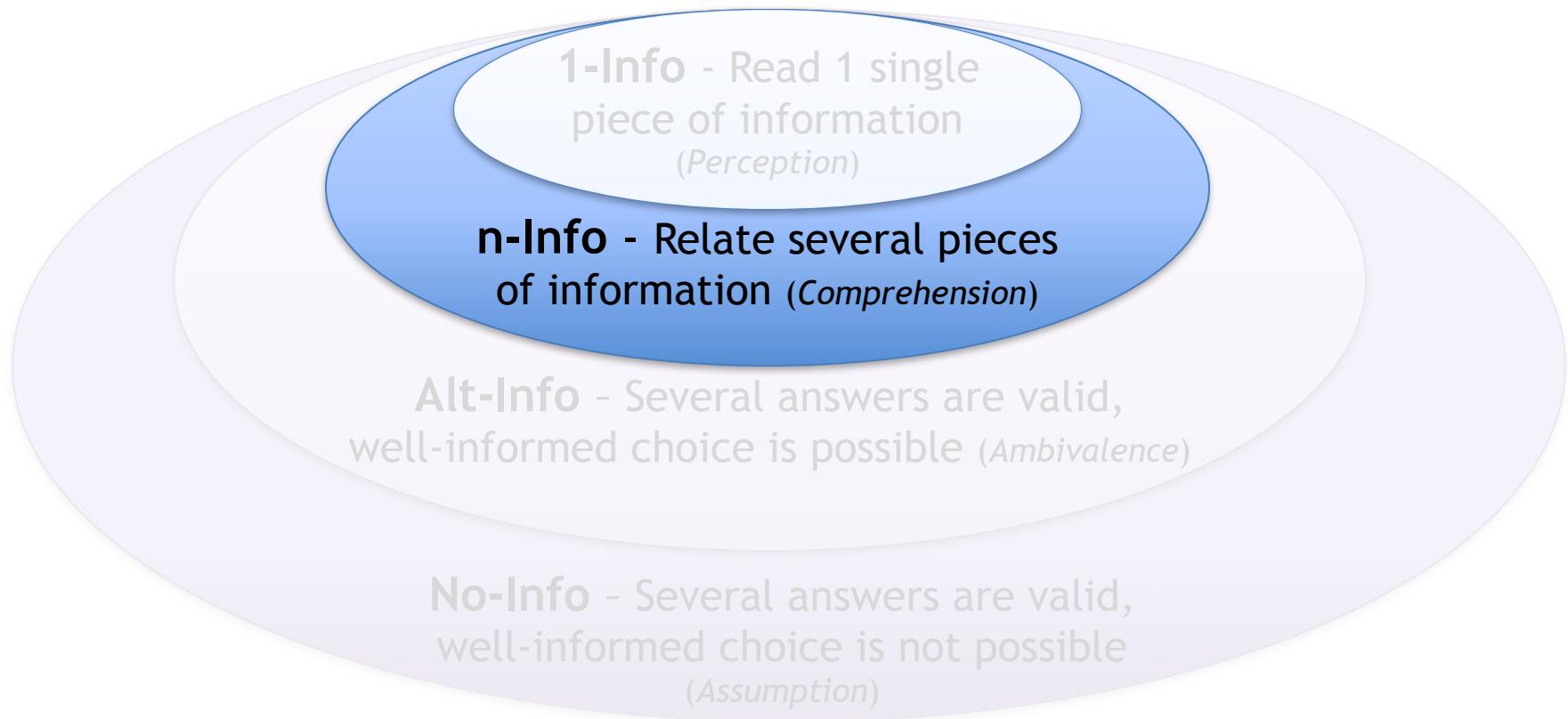
Alt-Info - Several answers are valid,
well-informed choice is possible
(*Ambivalence*)

No-Info - Several answers are valid,
well-informed choice is not possible
(*Assumption*)

Differentiate Levels of Complexity

“Which week has the most fish?”

“Is the number of video samples constant over the weeks?”



Differentiate Levels of Complexity

“What is the fish abundance for week 45?”

“Which period of the year has the highest fish abundance?”

1-Info - Read 1 single piece of information
(Perception)

n-Info - Relate several pieces of information
(Comprehension)

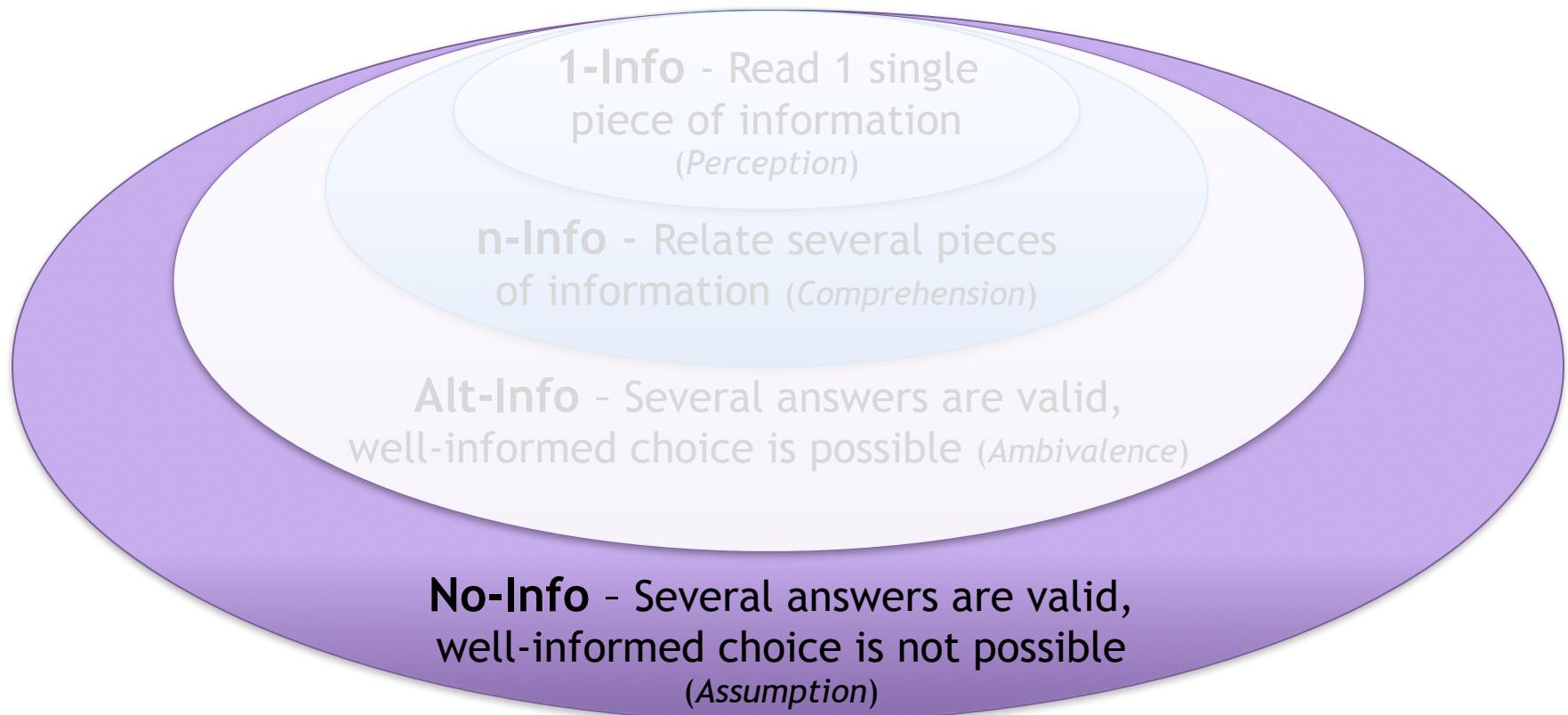
Alt-Info - Several answers are valid, well-informed choice is possible
(Ambivalence)

No-Info - Several answers are valid, well-informed choice is not possible
(Assumption)

Differentiate Levels of Complexity

“Is there a correlation in fish abundance for species 9, 26 and 27?”

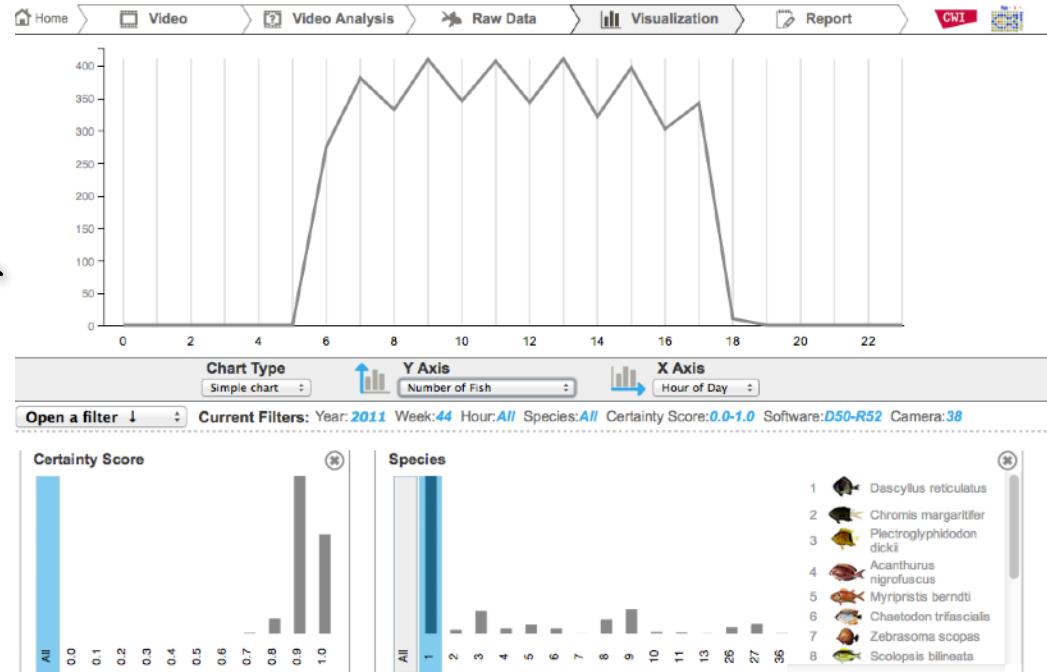
“Is the abundance of species 2 lower than species 1 because video analysis does not correctly detect species 2?”



Differentiate Layout & Interaction Design

What are the issues with the layout design?

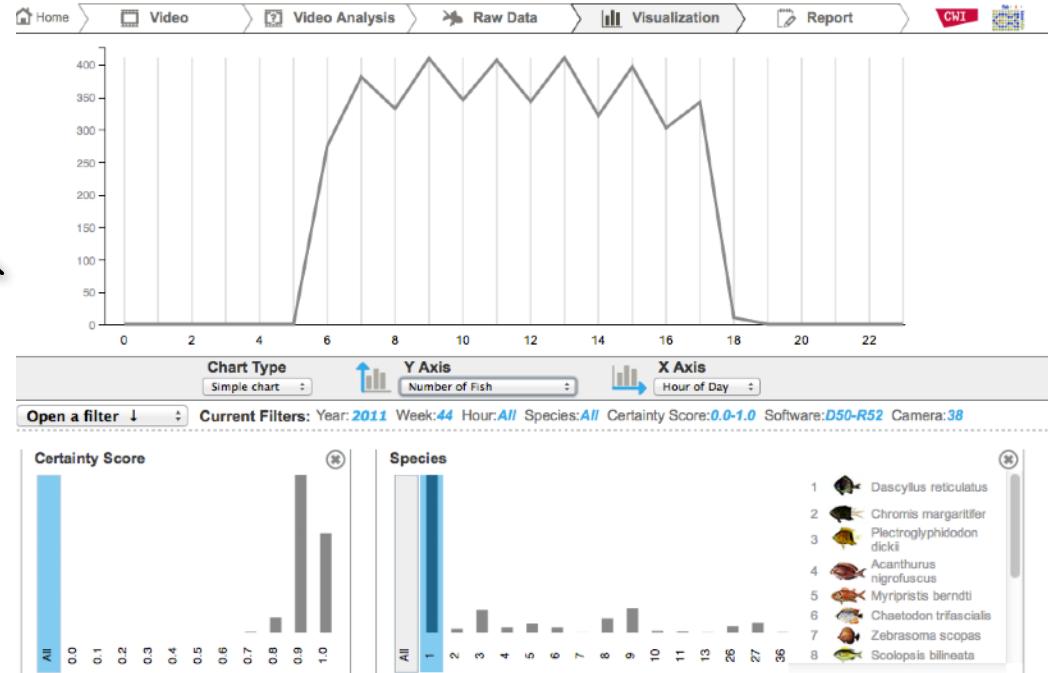
..or with the interaction design?



Differentiate Layout & Interaction Design

What are the issues with the layout design?

..or with the interaction design?



Probes (predefined states of the interface)
expose users to particular layouts prior to interaction

Differentiate Layout & Interaction Design

What are the issues with the layout design?

..or with the interaction design?



2 more groups of tasks:
No Interaction (no-IX) - Interaction Needed (IX)

20 tasks asked to 10 ecologists using 3 probes

User indicated their confidence in their answers
(5-Likert scale - *Very High* to *Very Low* confidence)

Low confidence can be the right level of confidence
which reflects the presence of uncertainty

20 tasks asked to 10 ecologists using 3 probes

User indicated their confidence in their answers
(5-Likert scale - *Very High* to *Very Low* confidence)

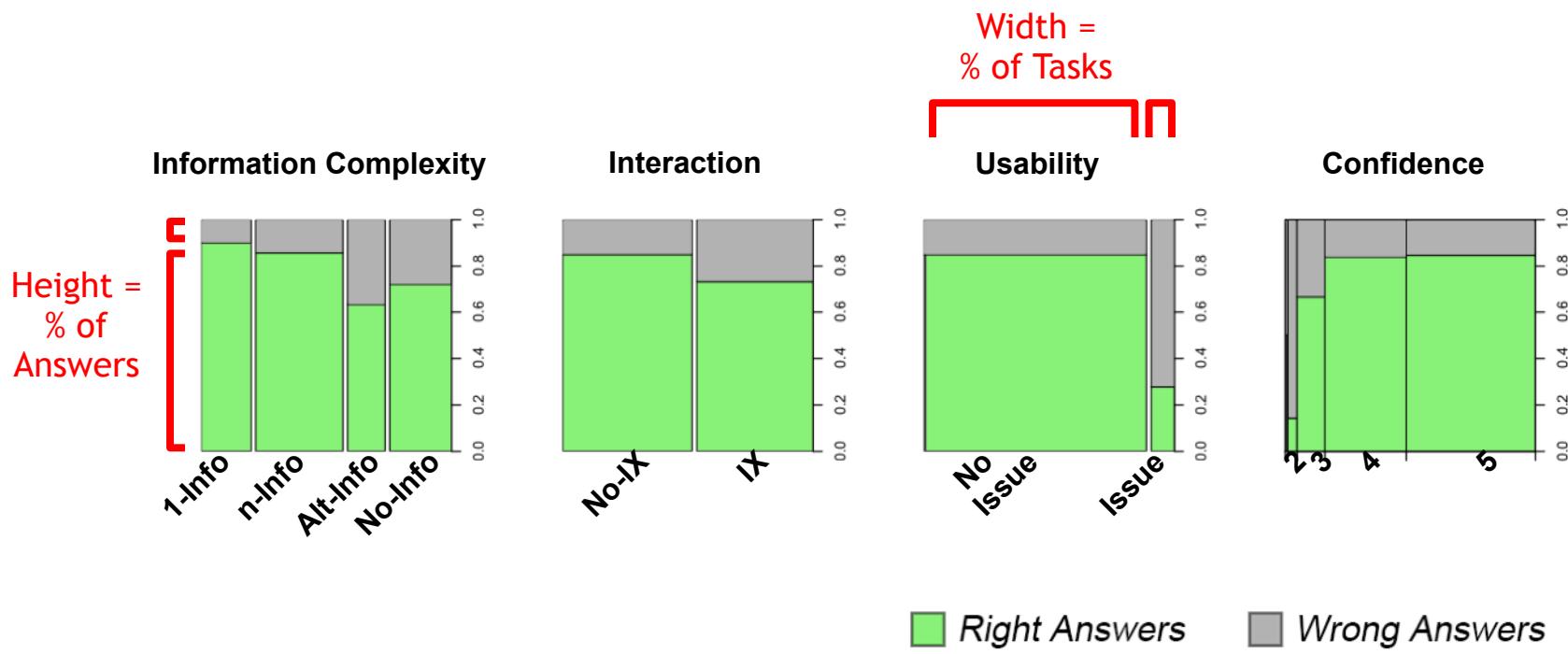
Low confidence can be the right level of confidence
which reflects the presence of uncertainty

We analyzed the relationships between **Confidence Levels**,
Task Complexity, **Answer Correctness** and **Usability Issues**



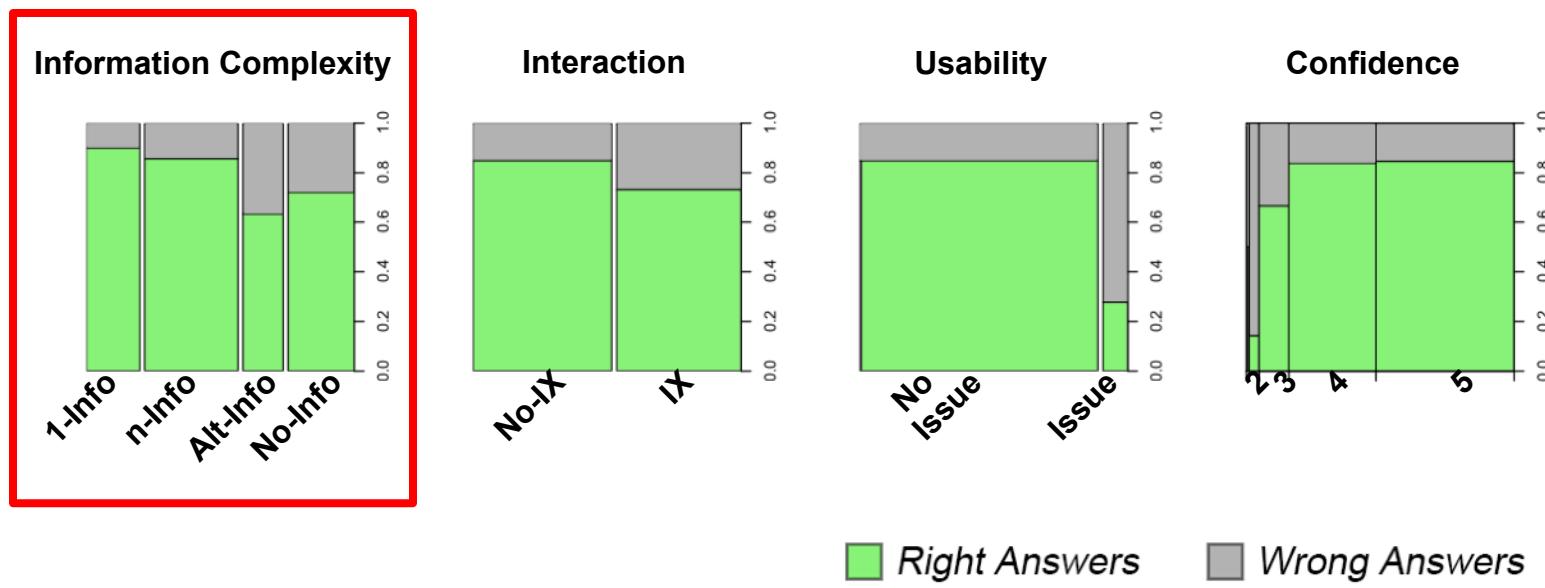
Experiment Results

Answer Correctness



Answer Correctness

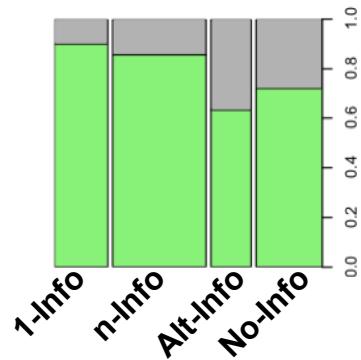
Answer Correctness decreased with
Information Complexity



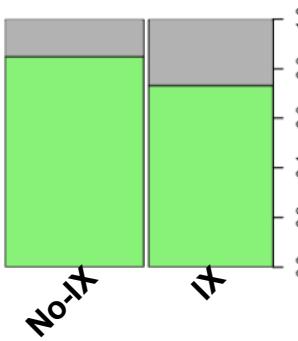
Answer Correctness

Answer Correctness decreased with
Interaction & Usability Features

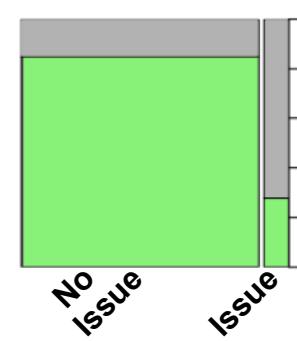
Information Complexity



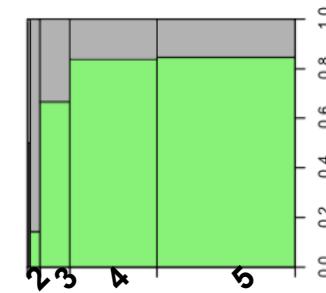
Interaction



Usability



Confidence

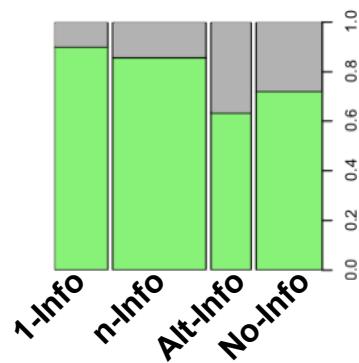


Right Answers Wrong Answers

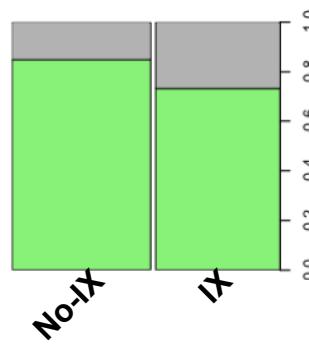
Answer Correctness

Answer Correctness decreased with Confidence

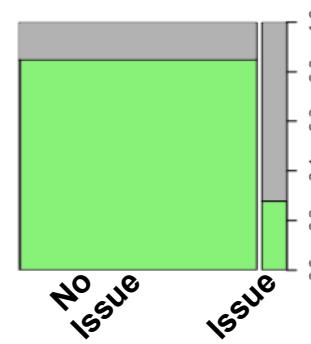
Information Complexity



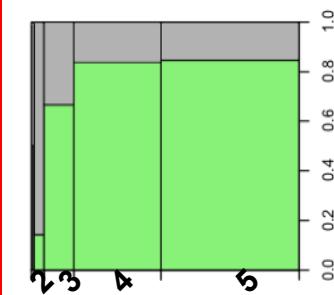
Interaction



Usability



Confidence

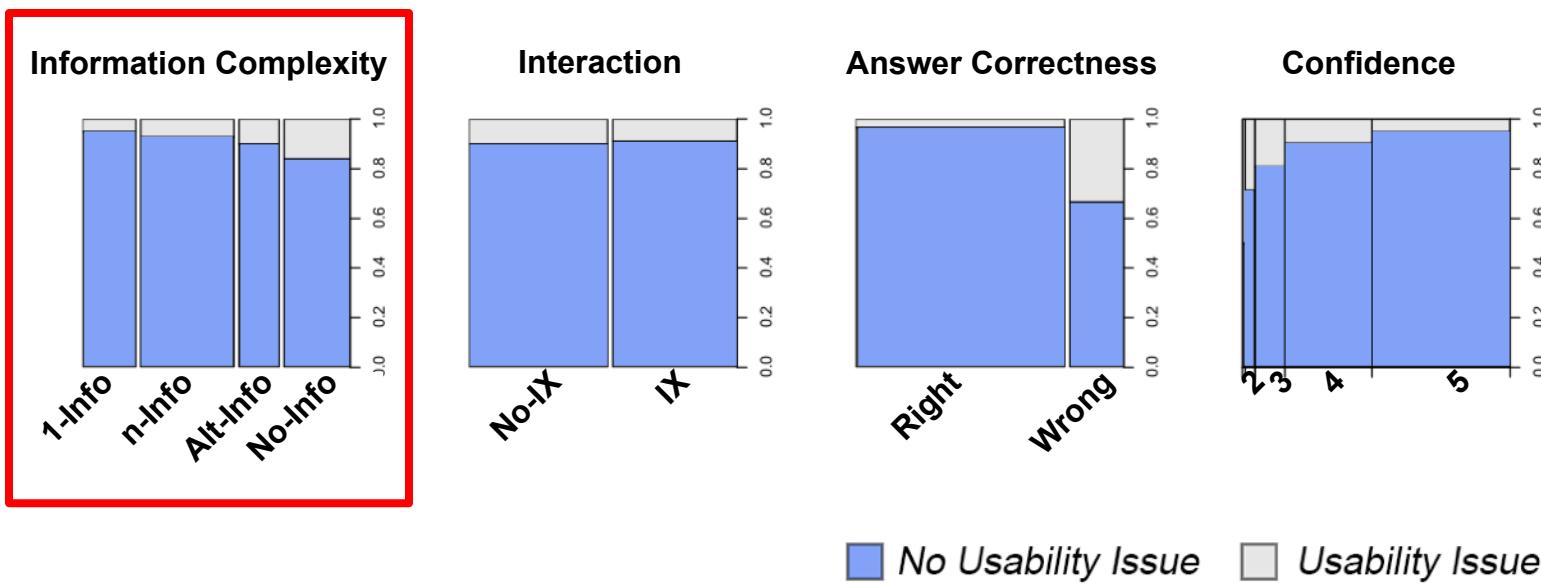


Right Answers

Wrong Answers

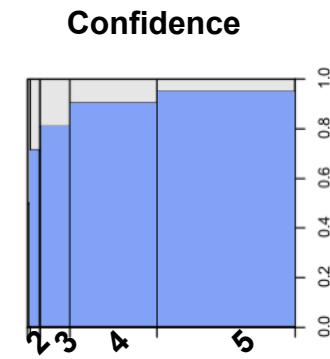
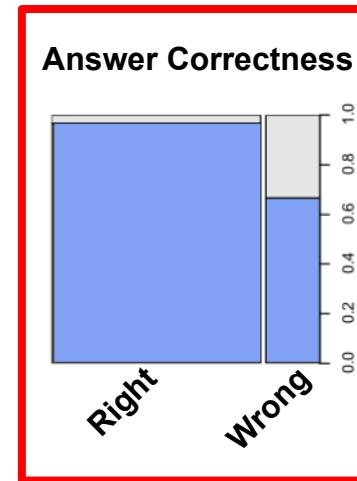
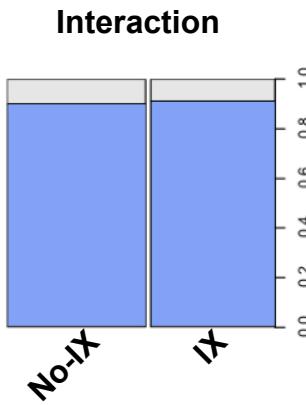
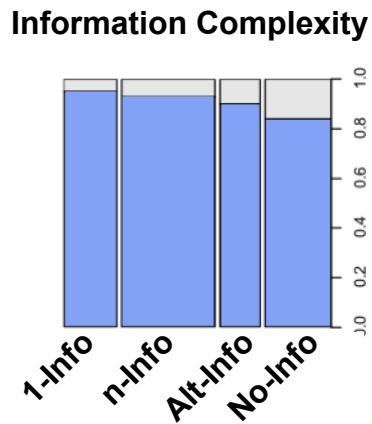
Usability Issues

Usability Issues increased with
Information Complexity



Usability Issues

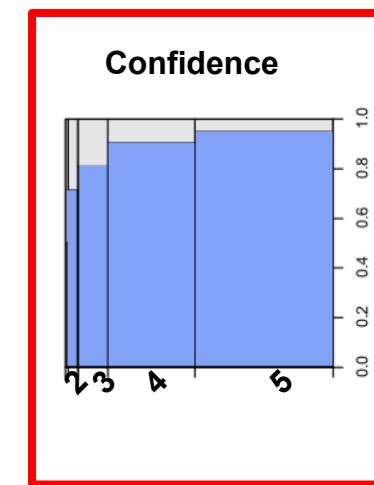
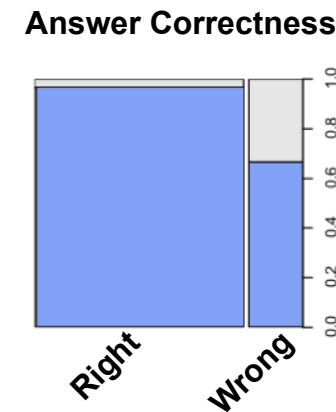
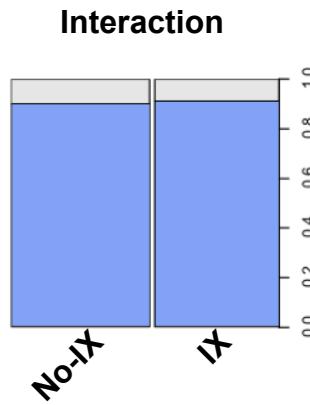
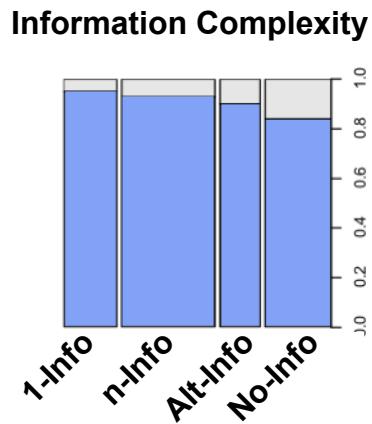
Usability Issues increased with
Wrong Answers



█ No Usability Issue █ Usability Issue

Usability Issues

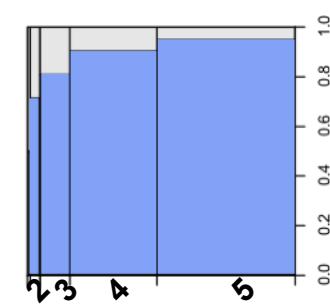
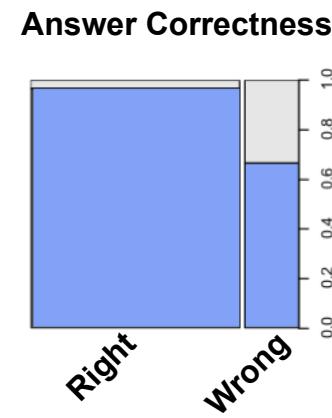
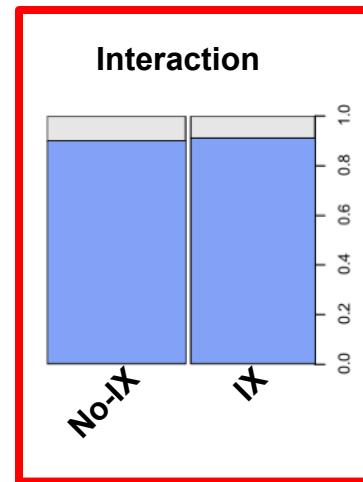
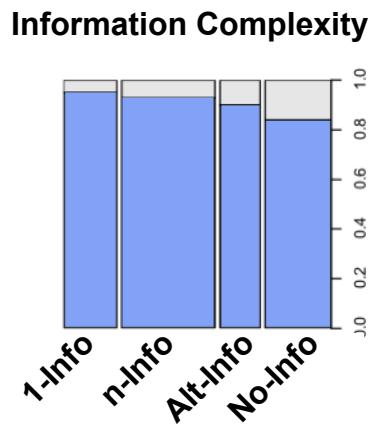
Usability Issues increased when
User Confidence decreased



█ No Usability Issue █ Usability Issue

Usability Issues

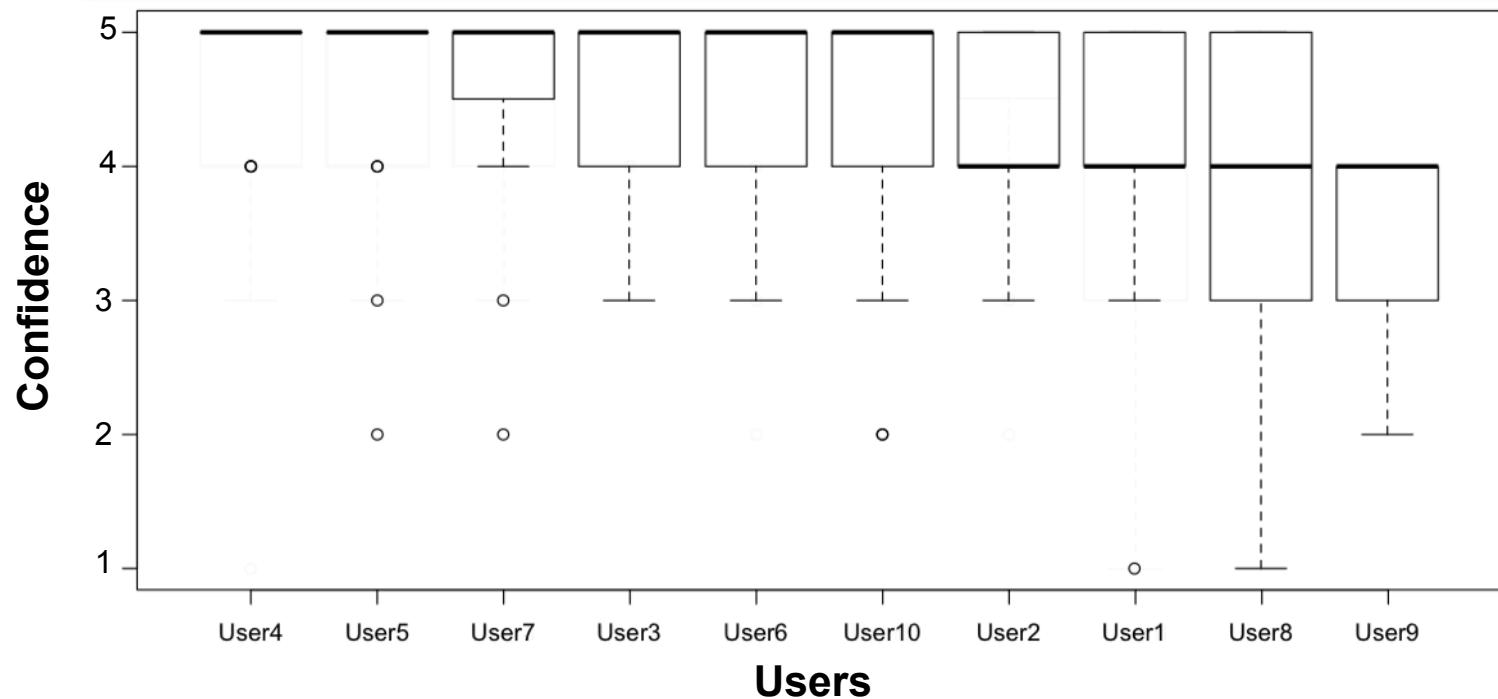
Usability Issues concerned both
layout & interactions



█ No Usability Issue █ Usability Issue

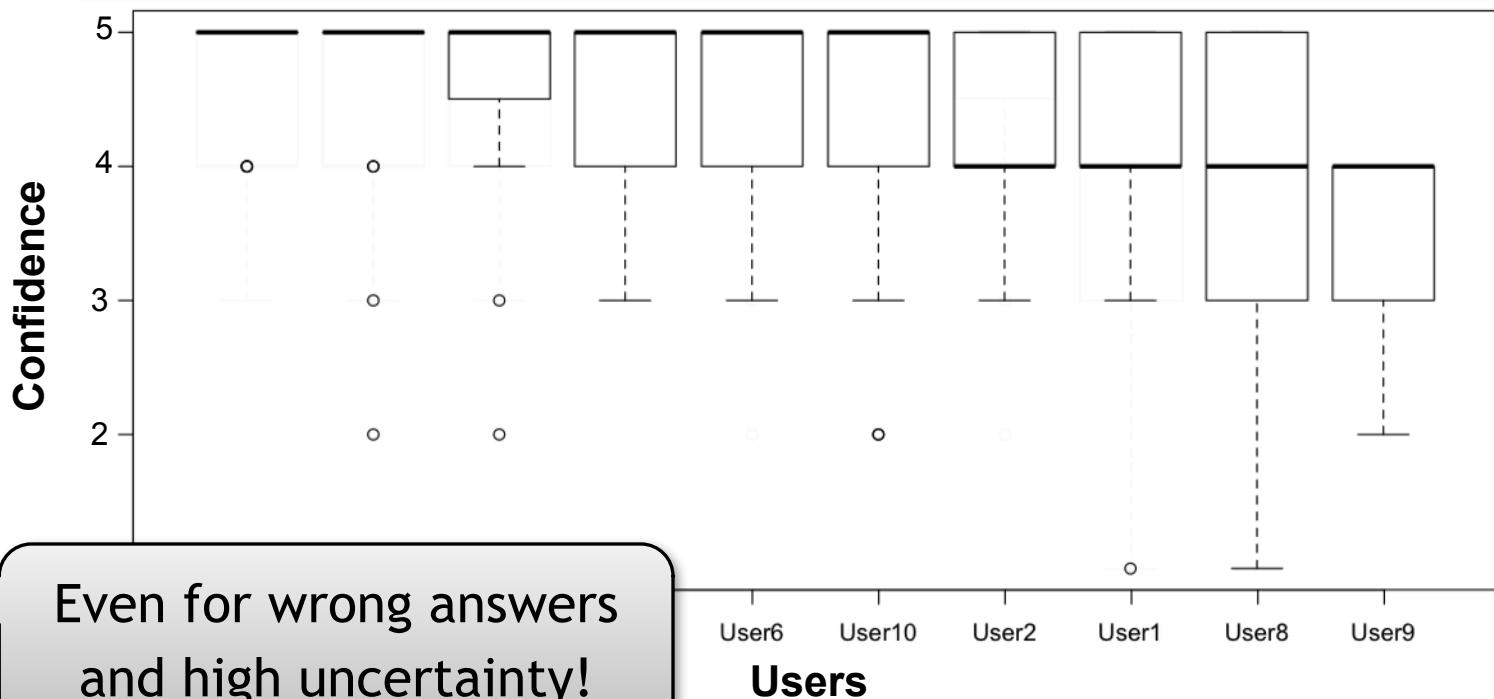
User Confidence

Users are generally highly confident



User Confidence

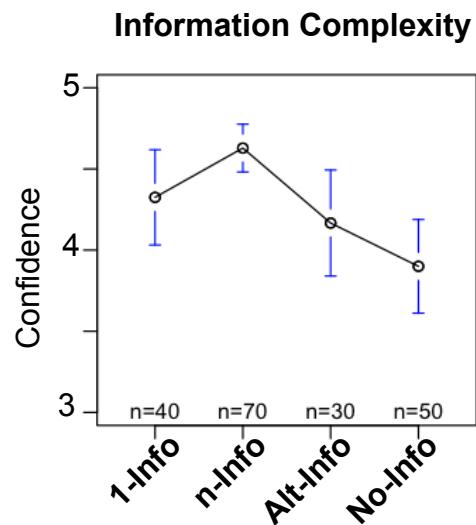
Users are generally highly confident



Even for wrong answers
and high uncertainty!

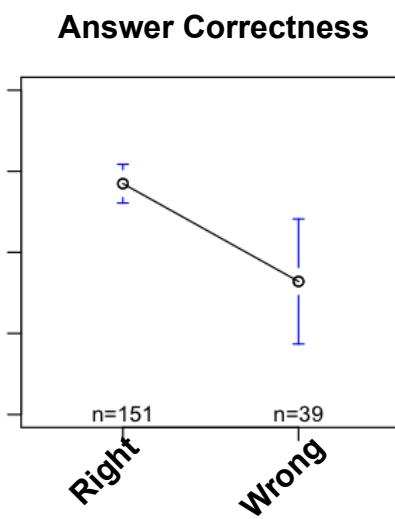
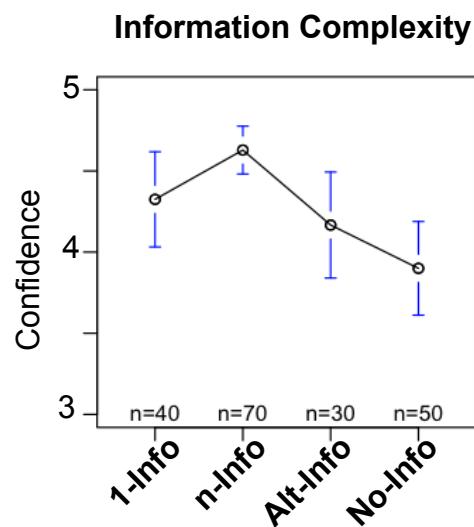
User Confidence

Confidence decreased with Information Complexity



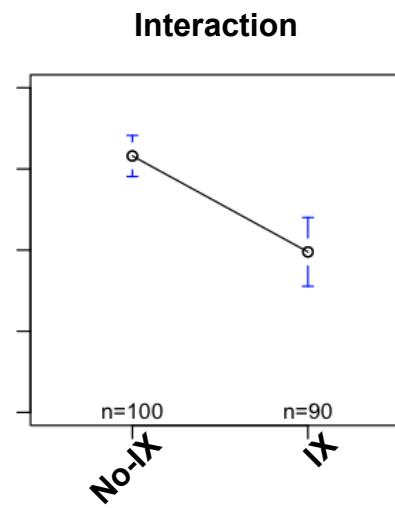
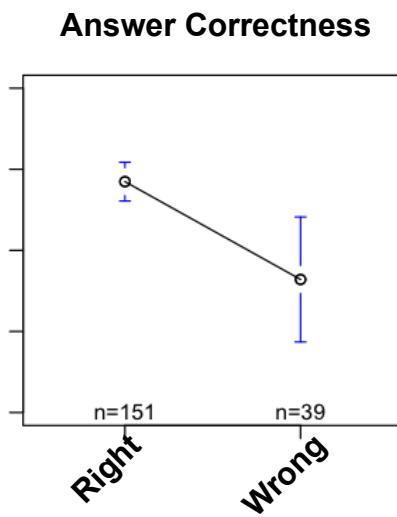
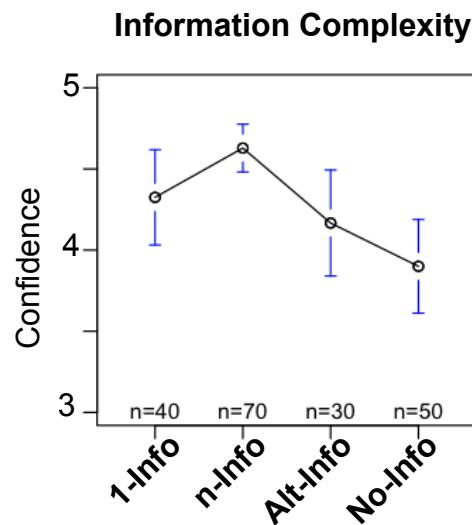
User Confidence

Confidence decreased with Information Complexity,
Answer Correctness



User Confidence

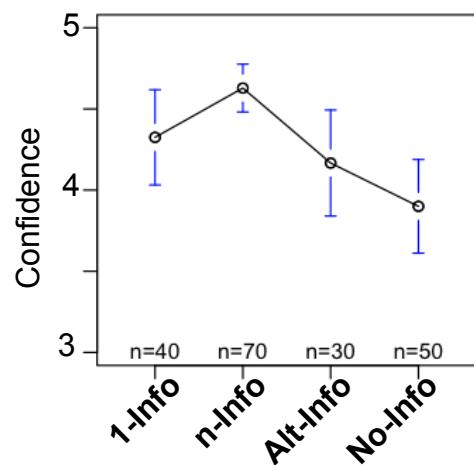
Confidence decreased with Information Complexity,
Answer Correctness, Interaction



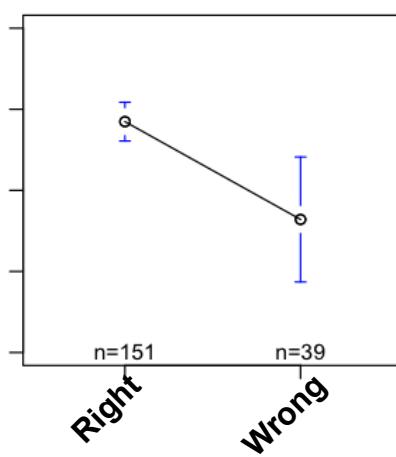
User Confidence

Confidence decreased with Information Complexity, Answer Correctness, Interaction and Usability Issues

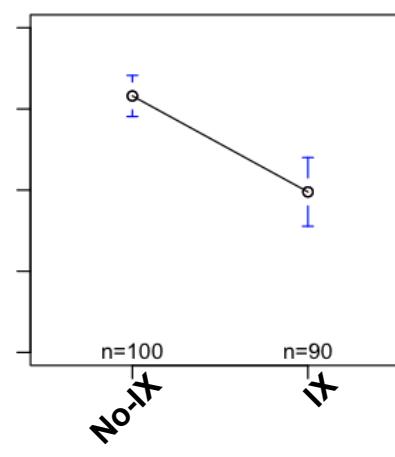
Information Complexity



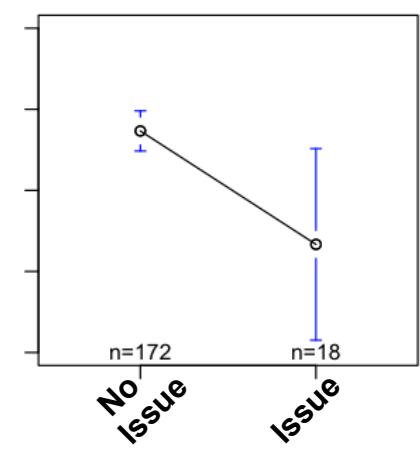
Answer Correctness



Interaction

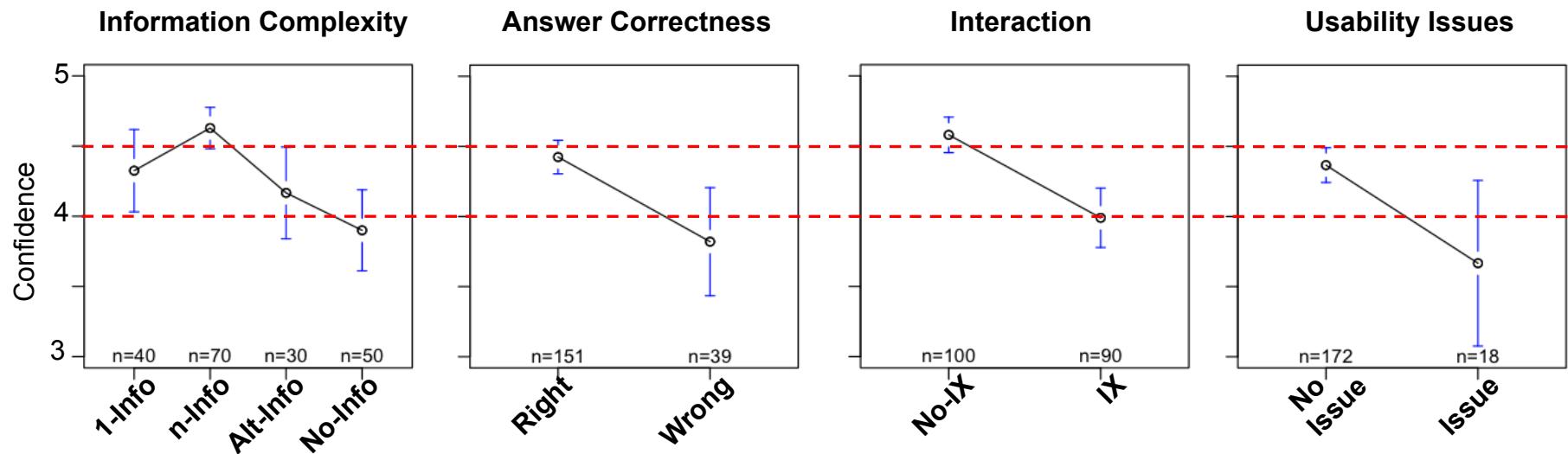


Usability Issues



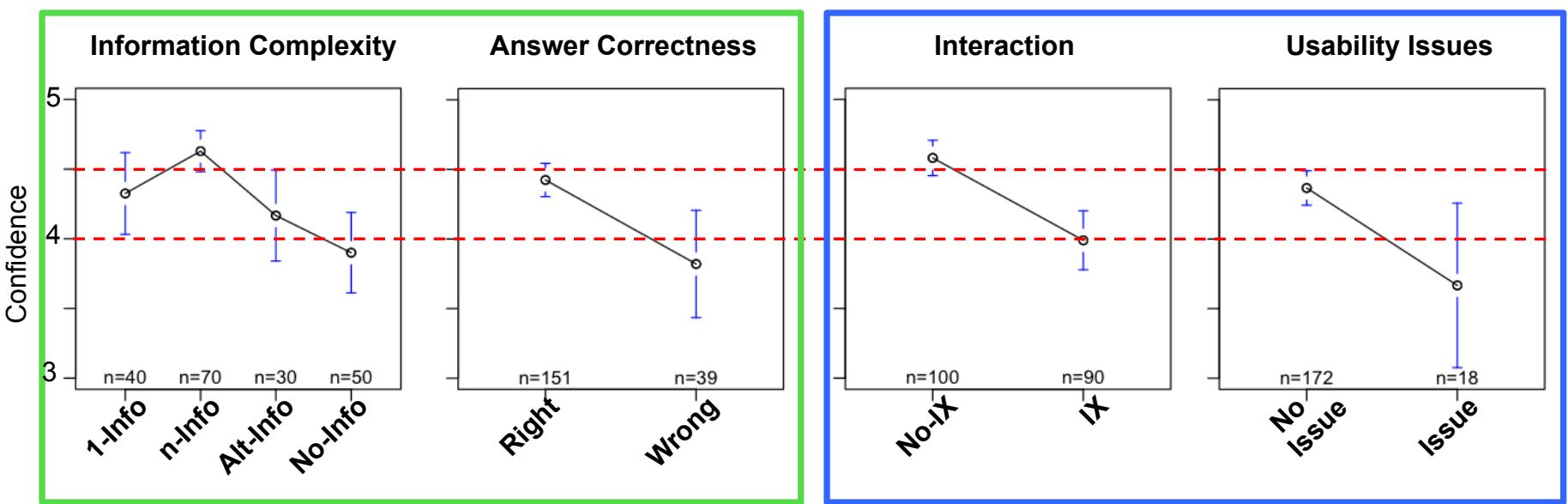
User Confidence

Confidence decreased with Information Complexity, Answer Correctness, Interaction and Usability Issues



User Confidence

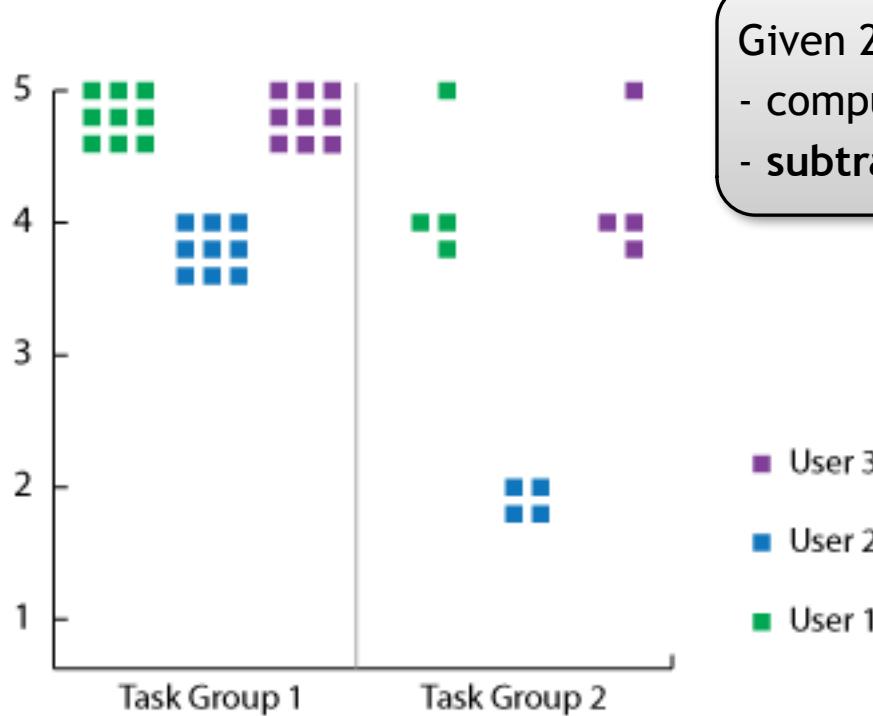
Did information & usability have a similar impact on user confidence?



Information Features Vs Usability Features

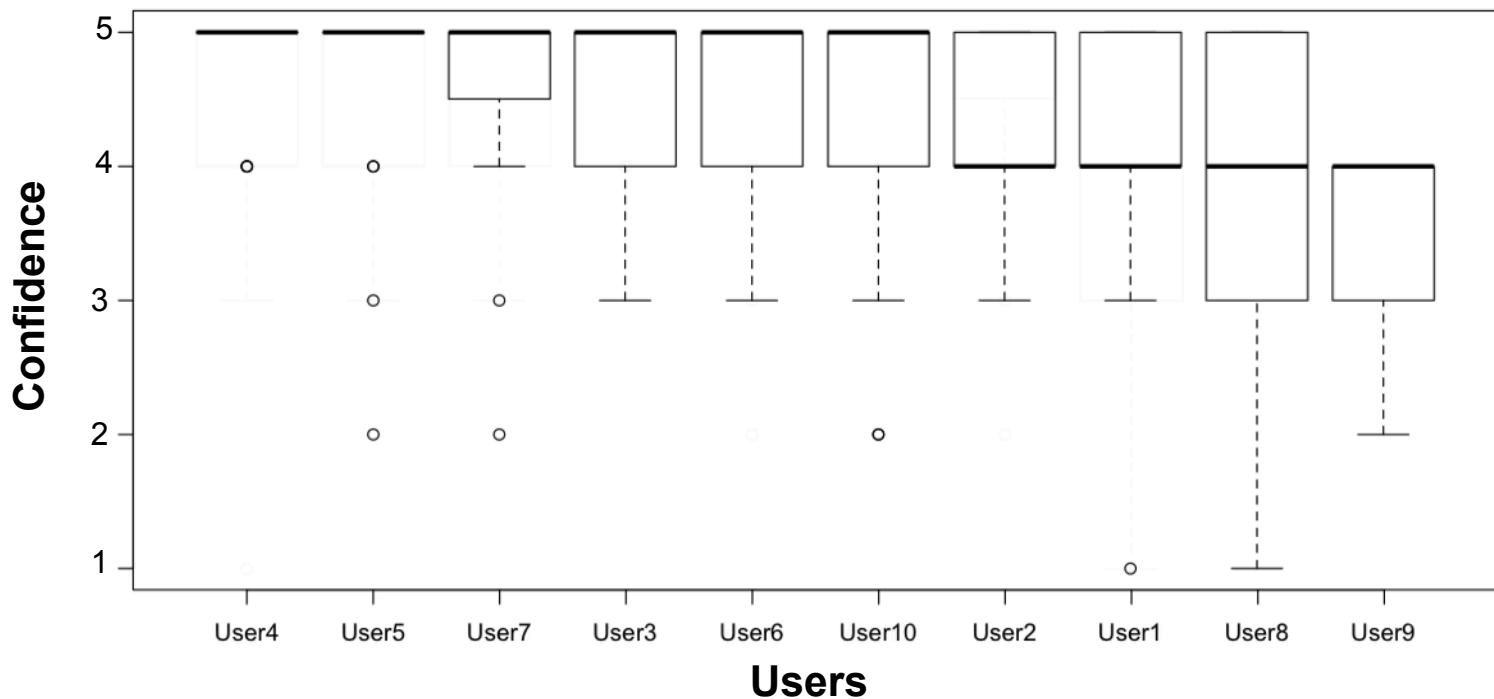
Confidence Drift Analysis

How much each user's confidence changed between groups of tasks?



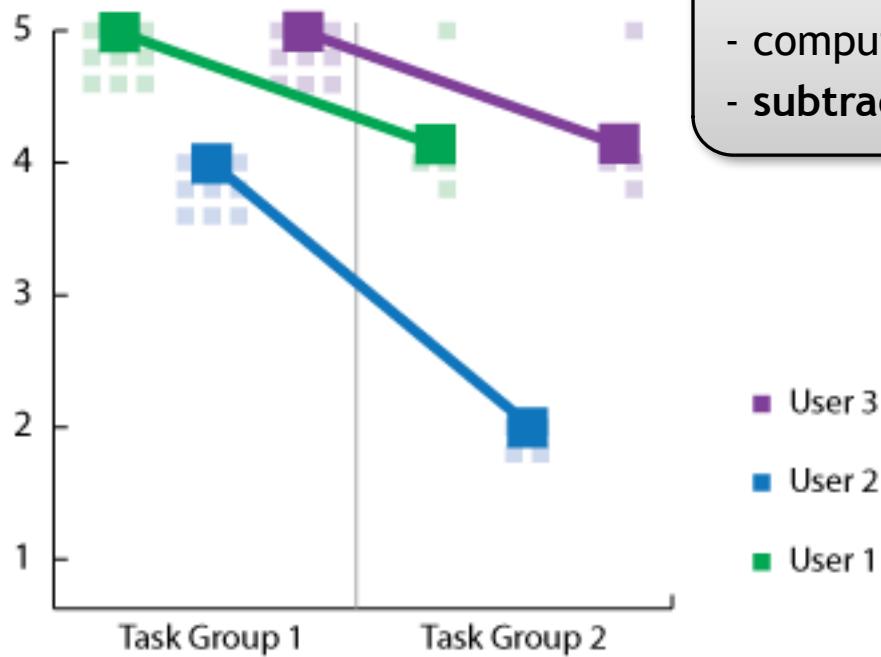
Given 2 groups of task, for each user separately:
- compute the **mean confidence** for each group
- subtract the 2 means to obtain the confidence drifts

Users' Default Level of Confidence



Confidence Drift Analysis

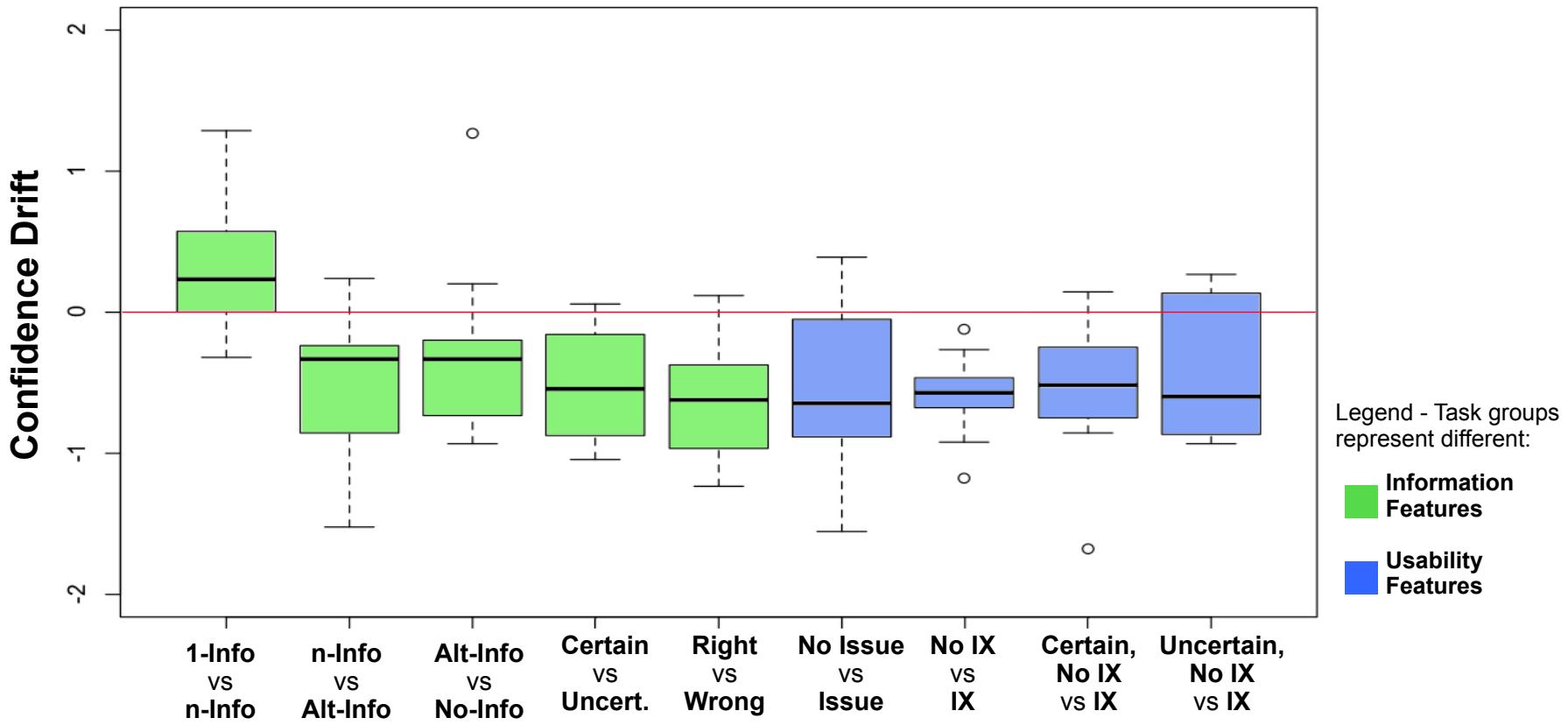
How much each user's confidence changed between groups of tasks?



Given 2 groups of task, for each user separately:

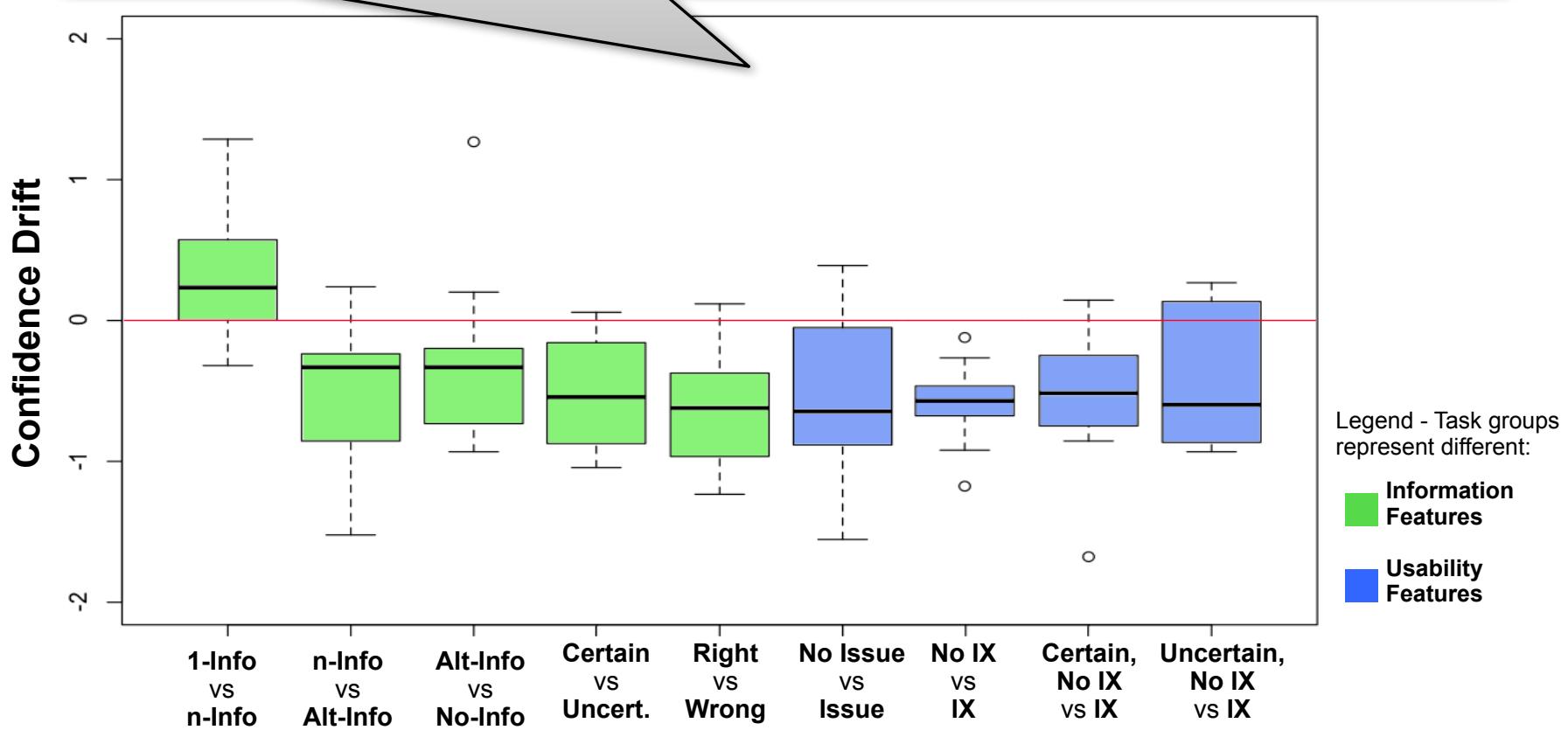
- compute the mean confidence for each group
- subtract the 2 means to obtain the confidence drifts

Confidence Drift Analysis



Confidence Drift Analysis

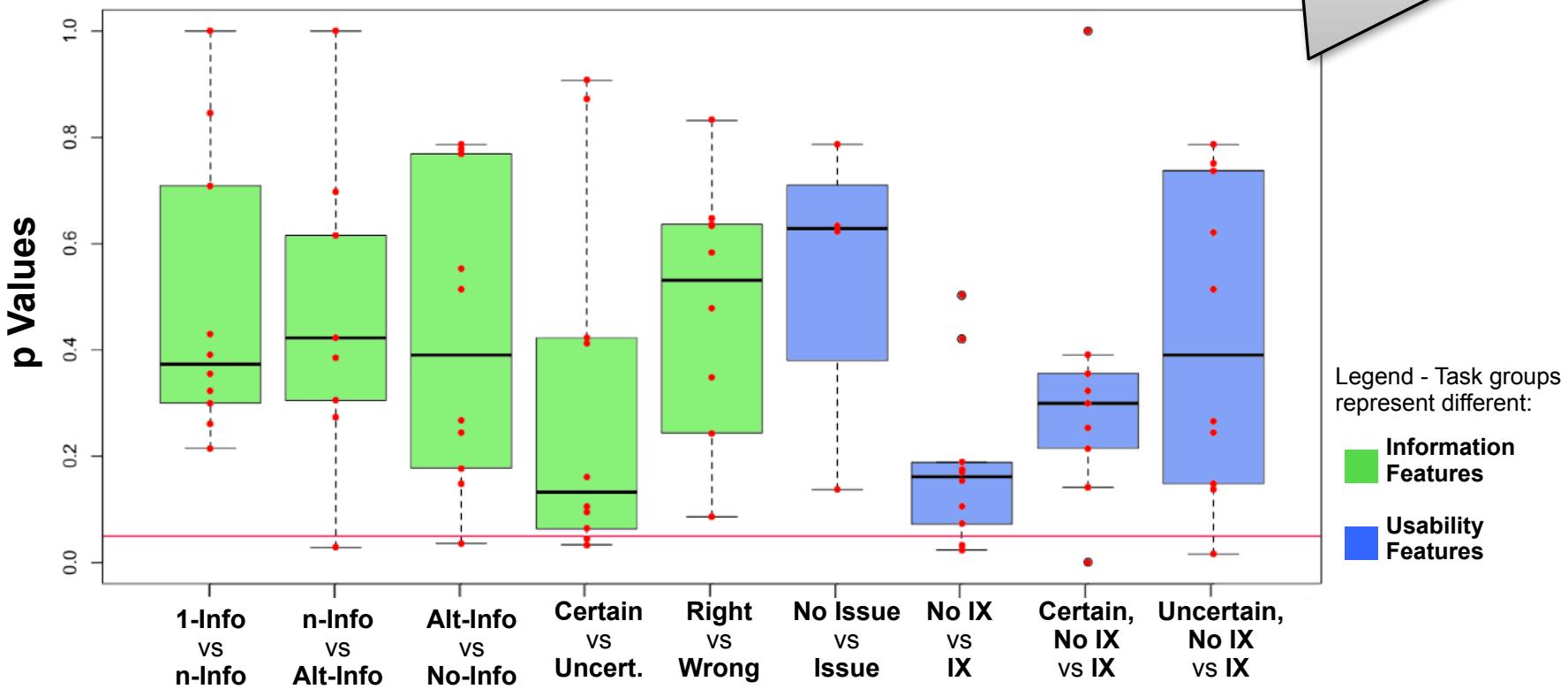
Our results show similar impacts
of Uncertainty Features and Usability Features on User Confidence



Confidence Drift Analysis

But our results are
not statistically significant

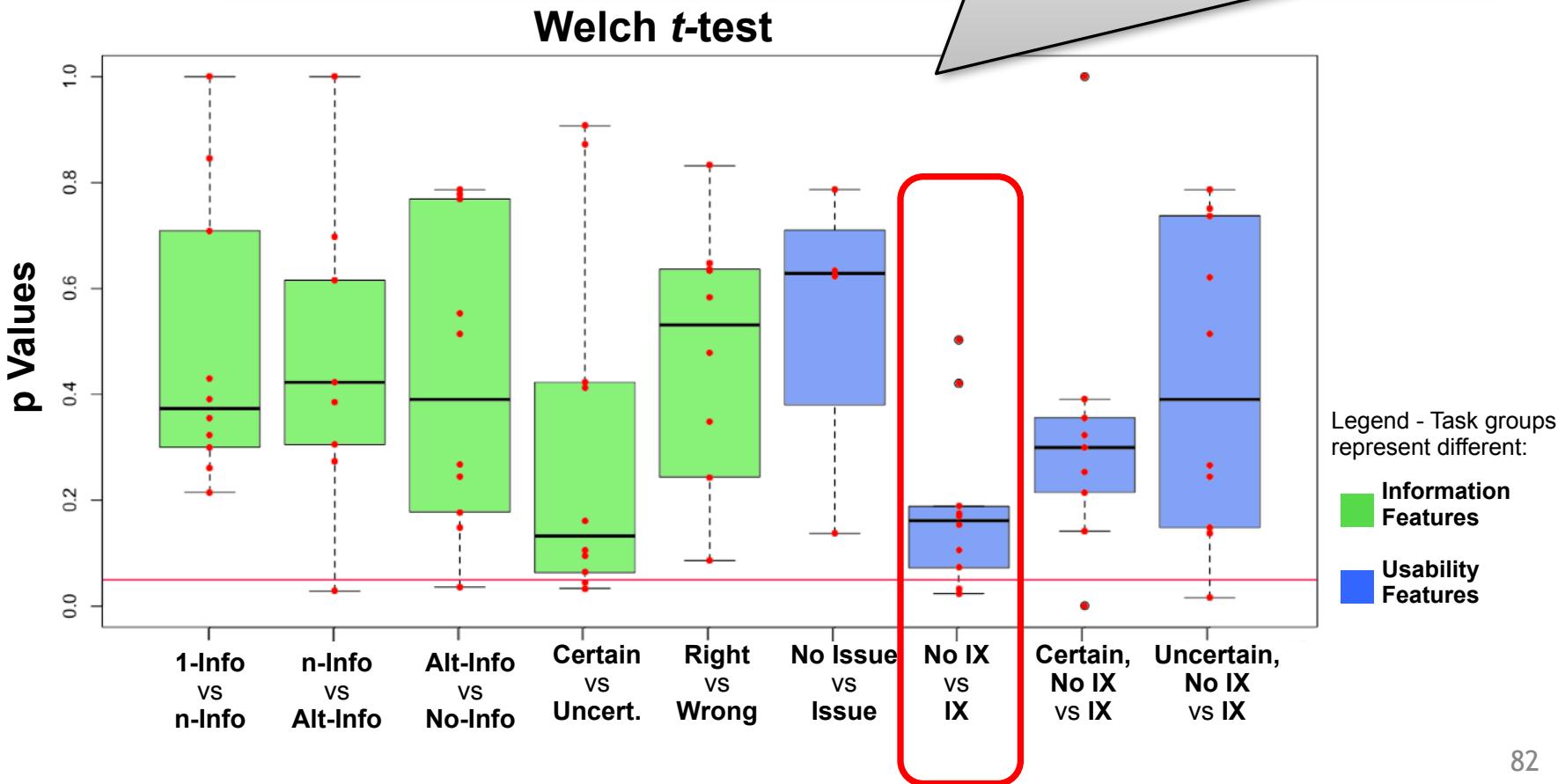
Welch *t*-test



Confidence Drift Analysis

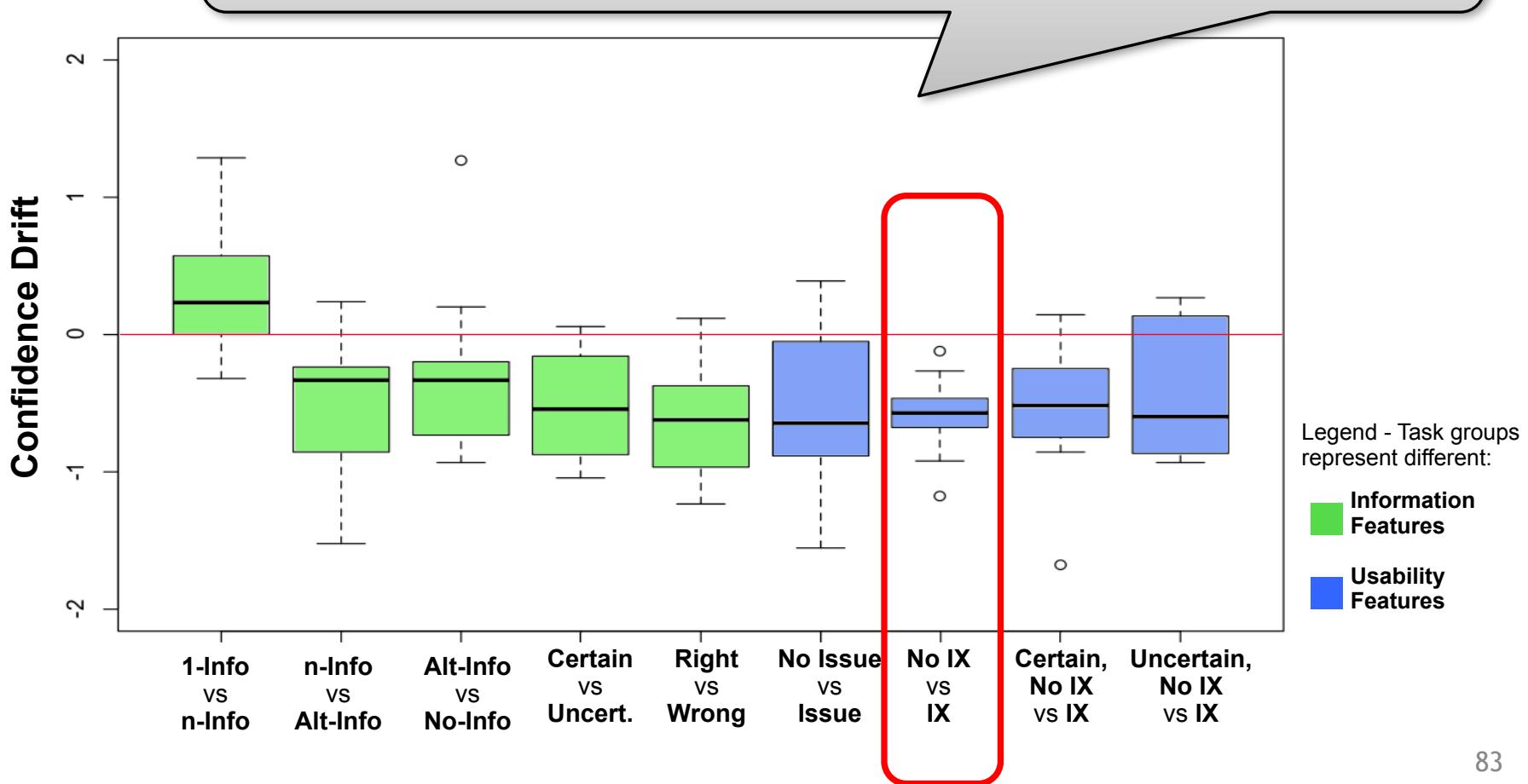
Yet the most significant impact is observed
when users interact with the interface

Welch t -test

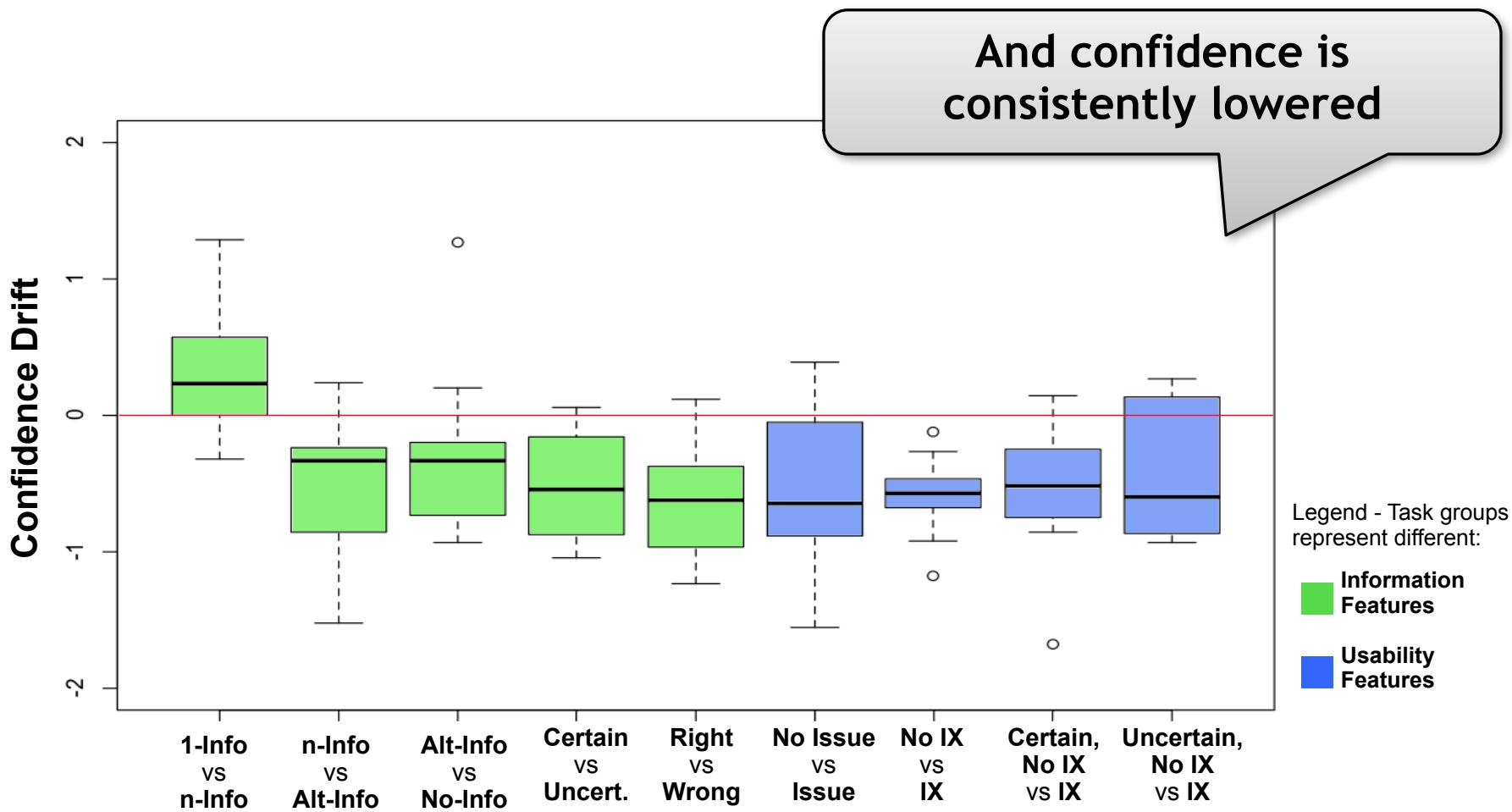


Confidence Drift Analysis

Yet the most significant impact is observed
when users interact with the interface

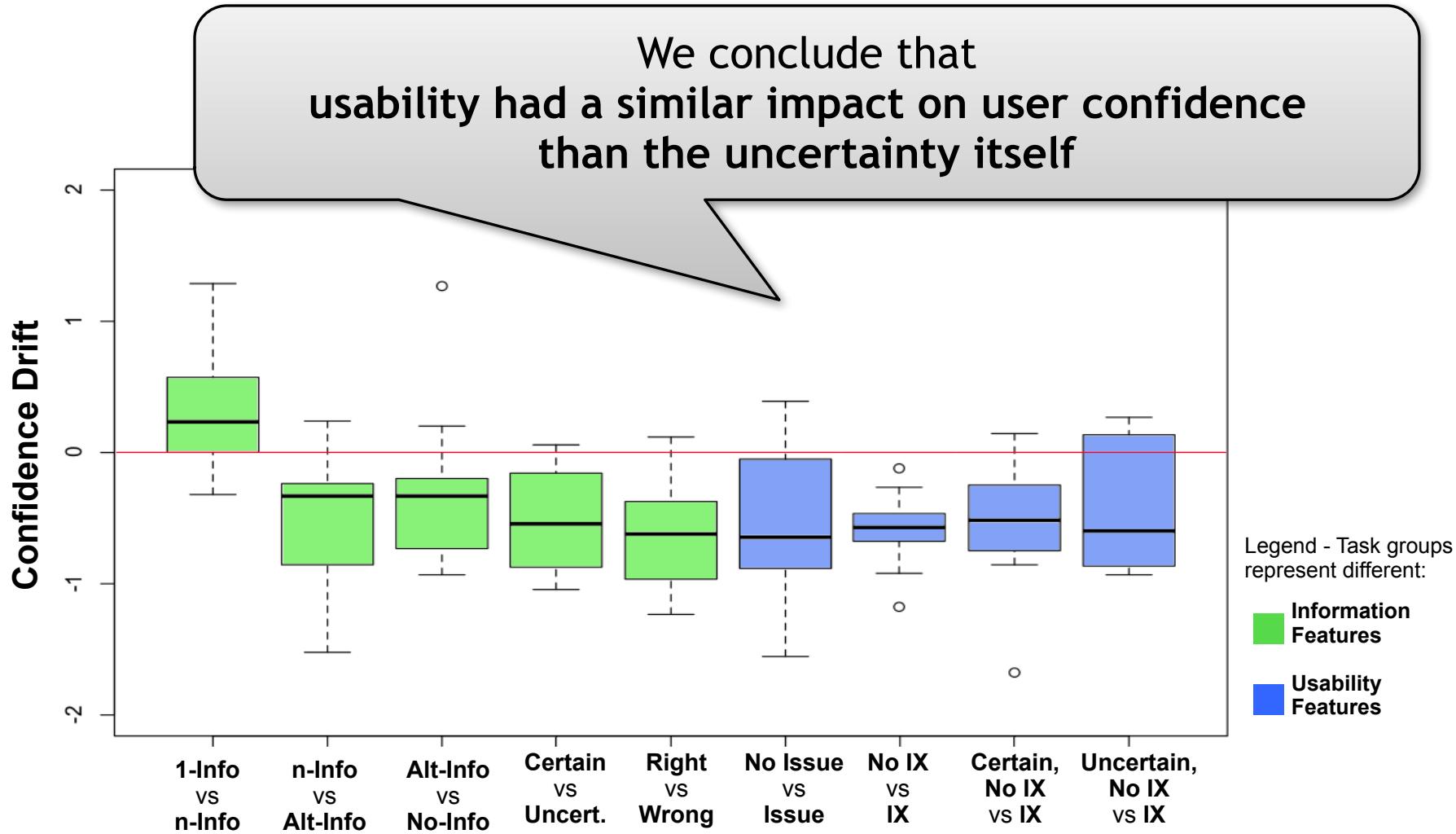


Confidence Drift Analysis



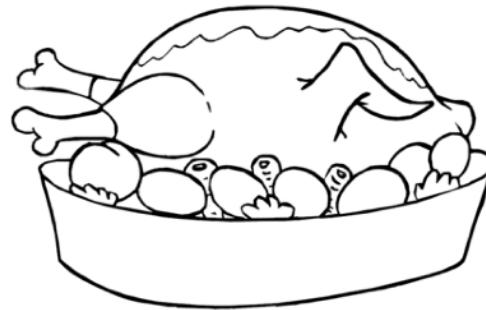
Confidence Drift Analysis

We conclude that
usability had a similar impact on user confidence
than the uncertainty itself



Conclusion

**Our design with simple graphs in multiple views
was easy to use and welcomed
but did not fully support user awareness of uncertainty**



Discussion

Design Issues

Usability had a similar impact on user confidence than the uncertainty itself

Design Issues

Usability had a similar impact on user confidence than the uncertainty itself

Computer Vision uncertainty is unfamiliar, and not anticipated

Design Issues

Usability had a similar impact on user confidence than the uncertainty itself

Computer Vision uncertainty is unfamiliar, and not anticipated

Simple graph may not signal underlying uncertainty

Design Issues

Usability had a similar impact on user confidence than the uncertainty itself

Computer Vision uncertainty is unfamiliar, and not anticipated

Simple graph may not signal underlying uncertainty

Multiple views may yield memory loss and attention tunneling

[22] Wickens & Carswell, 1997: Information Processing

[23] Wang Bonaldo *et al.*, 2000: Guidelines for using multiple views in visualization

Design Issues

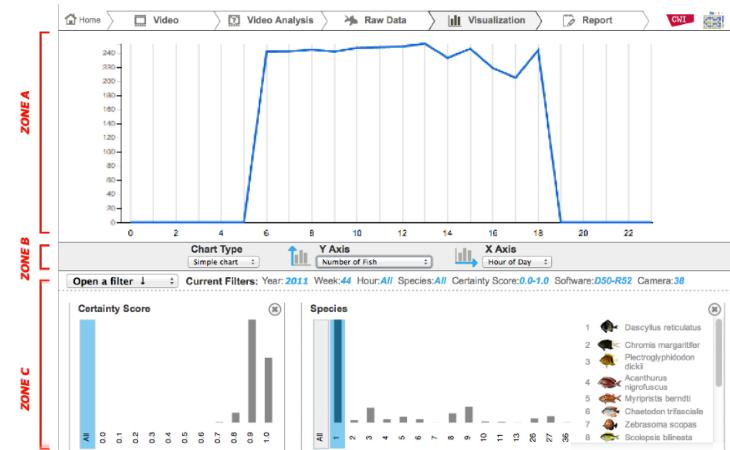
Usability had a similar impact on user confidence than the uncertainty itself

Computer Vision uncertainty is unfamiliar, and not anticipated

Simple graph may not signal underlying uncertainty

Multiple views may yield memory loss and attention tunneling

Increase views' consistency
(e.g., same Y axis)



Design Issues

Usability had a similar impact on user confidence than the uncertainty itself

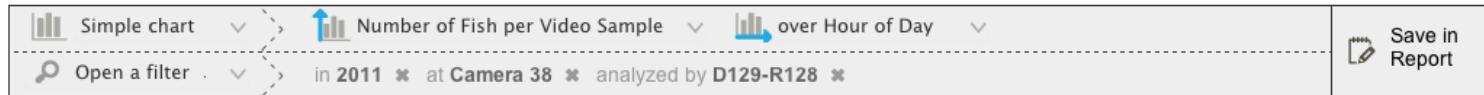
Computer Vision uncertainty is unfamiliar, and not anticipated

Simple graph may not signal underlying uncertainty

Multiple views may yield memory loss and attention tunneling

Increase views' consistency
(e.g., same Y axis)

Offload working memory
(e.g., delete details, use natural language)



Design Issues

Usability had a similar impact on user confidence than the uncertainty itself

Computer Vision uncertainty is unfamiliar, and not anticipated

Simple graph may not signal underlying uncertainty

Multiple views may yield memory loss and attention tunneling

Increase views' consistency
(e.g., same Y axis)

Offload working memory
(e.g., delete details, use natural language)

Increase the saliency of uncertainty (e.g., push that information)

Users' Over-Confidence

**Users were generally highly confident
even when uncertainty was high or answers were wrong**

Users' Over-Confidence

Users were generally highly confident even when uncertainty was high or answers were wrong

Possibly due to the presence of observers, or to cultural factors

Users' Over-Confidence

Users were generally highly confident even when uncertainty was high or answers were wrong

Possibly due to the presence of observers, or to cultural factors

User studies should give incentive to express low confidence
(10-Likert scale? Bonus-Malus?)



Overlooked Numbers of Samples

**Sampling size issues were overlooked
yet they are common in ecology research**

Overlooked Numbers of Samples

**Sampling size issues were overlooked
yet they are common in ecology research**

Possibly because widgets used Number of Fish by default
thus overshadowing Average Fish per Video

Overlooked Numbers of Samples

**Sampling size issues were overlooked
yet they are common in ecology research**

Possibly because widgets used Number of Fish by default
thus overshadowing Average Fish per Video

Possibly due to a different terminology
(e.g., *number of videos Vs sampling size*)

Overlooked Numbers of Samples

Sampling size issues were overlooked
yet they are common in ecology research

Possibly because widgets used Number of Fish by default
thus overshadowing Average Fish per Video

Possibly due to a different terminology
(e.g., *number of videos Vs sampling size*)

Use “Mean Count per Video Sample”
as the default metric of fish abundance

Wrap-Up

Usability had a similar impact on user confidence than the uncertainty itself

Users were over-confident

Sampling size issues were overlooked

**Our code is open-source! Write me to get it:
e.m.a.l.beauxis@hva.nl**

EMMA BEAUXIS-AUSSALET

e.m.a.l.beauxisaussalet@vu.nl

@emma_beauxis

QUESTION / DISCUSSION

