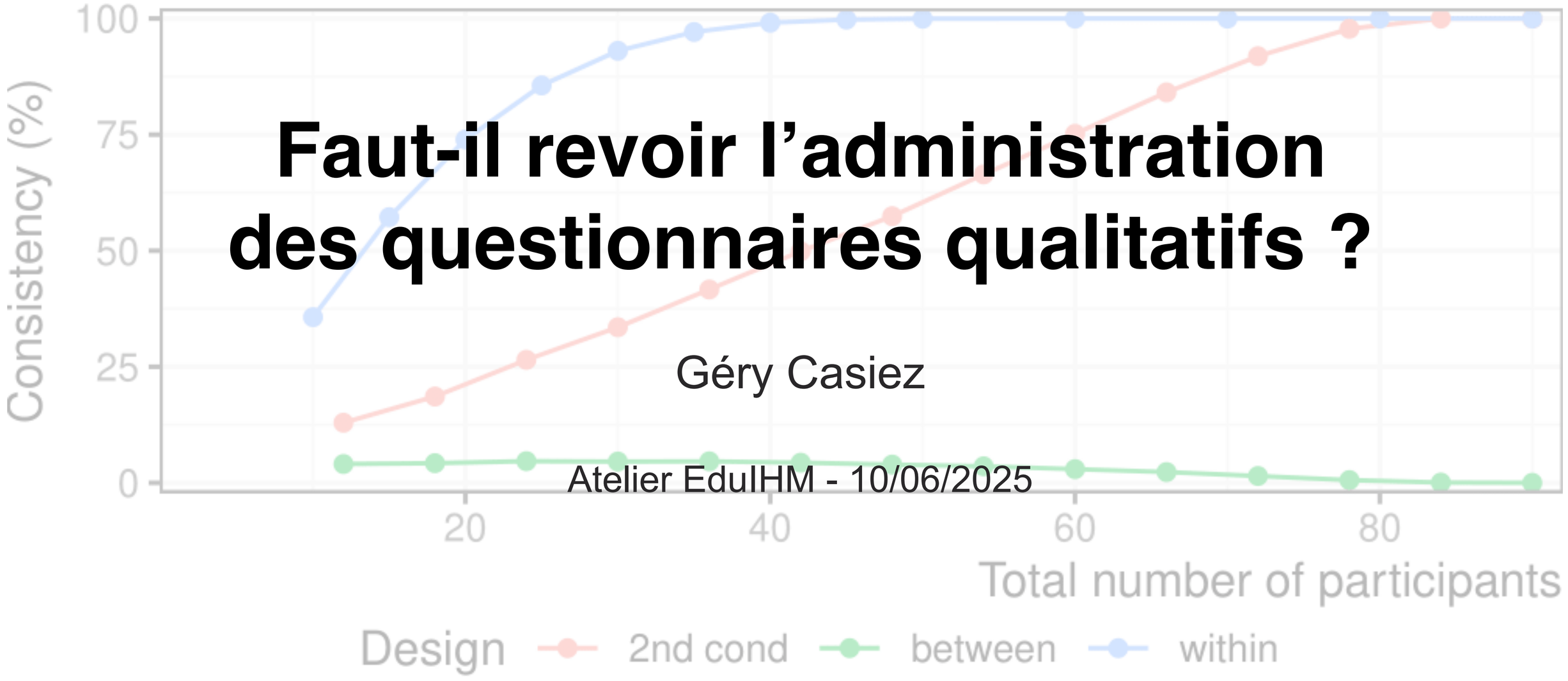
Faut-il revoir l'administration des questionnaires qualitatifs ?

Géry Casiez

Atelier EduIHM - 10/06/2025



Géry **Casiez**

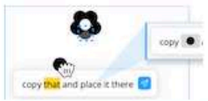
Professeur des Universités en informatique à l'Université de Lille
Professor of Computer Science at Univ. Lille
Junior member (2018 - 2023) of Institut Universitaire de France (IUF)
Adjunct Professor (2018 - 2028) at the University of Waterloo, Canada
Member of the Inria Loki research team
Member of CRIStAL

Inria, office B103
Parc scientifique de la Haute-Borne
40, Avenue Halley - Bât B - Park Plaza
59650 Villeneuve d'Ascq
+33 (0)3 59 35 87 87
gery (dot) casiez (at) univ-lille (dot) fr

CRIStAL
Bureau S1.20, ESPRIT, Cité Scientifique
+33 (0)3 28 77 85 44

I'm a Full Professor of Computer Science, Classe Exceptionnelle, at Université de Lille. My research is in the area of Human-Computer Interaction (HCI), carried in the Loki research team, created in partnership between the Inria centre at the University of Lille and the Université de Lille within the Joint Research Unit UMR 9189 CRIStAL. I teach in the CS department at IUT de Lille. My works have been published at ACM CHI, ACM UIST, IEEE VR or IHM, for which I've participated as a PC member or in the organization. In 2018, I have been appointed junior member of the Institut Universitaire de France, became Adjunct-Professor at the University of Waterloo and deputy director of CRIStAL. My research interests include input interaction (design of input devices, interaction techniques, study of transfer functions, latency measure and compensation, and noise reduction in interactive systems), VR and 3D interaction (study of embodiment and design of interaction techniques and input devices), interactive documents, haptics, multitouch interaction and BCI.

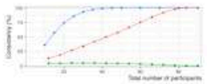
Selected publications



Masson, D., Malacria, S., Casiez, G. & Vogel, D. (2024). DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. *In proceedings of CHI'24, the 41st Conference on Human Factors in Computing Systems*, 1-16. ACM. AR: 1060/4028 (26%). HAL **Honorable mention award (top 5%)** Post on Medium

Masson, D., Malacria, S., Vogel, D., Lank, E. & Casiez, G. (2023). ChartDetective: Easy and Accurate Interactive Data Extraction from Complex Vector Charts. *In proceedings of CHI'23, the 40th Conference on Human Factors in Computing Systems*, 1-17. ACM. AR: 880/3182 (28%). HAL [Presentation] **Best paper award (top 1%)** Used for systematic literature reviews such as [Suponin et al.]

Masson, D., Malacria, S., Casiez, G. & Vogel, D. (2023). Charagraph: Interactive Generation of Charts for Realtime Annotation of Data-Rich Paragraphs. *In proceedings of CHI'23, the 40th Conference on Human Factors in Computing Systems*, 1-18. ACM. AR: 880/3182 (28%). HAL [Presentation]



Richard, G., Pietrzak, T., Argelaguet, F., Lécuyer, A. & Casiez, G. (2022). Within or Between? Comparing Experimental Designs for Virtual Embodiment Studies. *In proceedings of IEEE International Conference on Virtual Reality (IEEE VR)*, 186-195. IEEE. AR: 90/418 (22%). HAL [Presentation]

Questionnaires qualitatifs

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

System Usability Scale

	Pas du tout d'accord				Tout à fait d'accord
1. Je pense que je vais utiliser ce service fréquemment	1	2	3	4	5
2. Je trouve ce service inutilement complexe	1	2	3	4	5
3. Je pense que ce service est facile à utiliser	1	2	3	4	5
4.Je pense que j'aurai besoin de l'aide d'un technicien pour être capable d'utiliser ce service	1	2	3	4	5
5.J'ai trouvé que les différentes fonctions de ce service ont été bien intégrées	1	2	3	4	5
6. Je pense qu'il y a trop d'incohérence dans ce service.	1	2	3	4	5
7. J'imagine que la plupart des gens serait capable d'apprendre à utiliser ce services très rapidement.	1	2	3	4	5
8 J'ai trouvé ce service trés lourd à utiliser.	1	2	3	4	5
9. Je me sentais très en confiance en utilisant ce service.	1	2	3	4	5
10. J'ai besoin d'apprendre beaucoup de choses avant de pouvoir utiliser ce service.	1	2	3	4	5

Q1. I felt out of my body"

never	almost never	rarely	half of the time	often	most of the time	always
(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)

Q2. I felt as if my (real) *body* were drifting toward the virtual *body* or as if the virtual *body* were drifting toward my (real) *body*"

never	almost never	rarely	half of the time	often	most of the time	always
(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)

Q3. I felt as if the movements of the virtual *body* were influencing my own movements"

never	almost never	rarely	half of the time	often	most of the time	always
(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)

Q4. It felt as if my (real) *body* were turning into an 'avatar' *body*"

never	almost never	rarely	half of the time	often	most of the time	always
(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)

Q5. At some point it felt as if my real *body* was starting to take on the posture or shape of the virtual *body* that I saw"

never	almost never	rarely	half of the time	often	most of the time	always
(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)

Q6. I felt like I was wearing different clothes from when I came to the laboratory"

never	almost never	rarely	half of the time	often	most of the time	always
(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)

Q7. I felt as if my *body* had changed"

never	almost never	rarely	half of the time	often	most of the time	always
(-3)	(-2)	(-1)	(0)	(1)	(2)	(3)

Questionnaires qualitatifs

Tableau 1. Résumé des items du questionnaire d'acceptabilité

Dimensions	Items
Attente de performance	Le jeu *FunSpeech* va aider mon enfant à produire des sons. *FunSpeech* permettra à mon enfant de développer son langage.
Attente d'effort	Le jeu me semble facile à utiliser. Ce jeu me semble difficile à utiliser. (i.e., item inversé)
Influence sociale	Je pense qu'il est déconseillé d'utiliser une tablette tactile avec mon enfant. Mon entourage pense qu'il est bien d'utiliser ce jeu pour aider mon enfant.
Conditions facilitatrices	Je sais qui contacter si je rencontre des difficultés avec le jeu. Je pense disposer des compétences nécessaires pour utiliser le jeu avec mon enfant à domicile
Motivation hédonique	Utiliser *FunSpeech* sera une activité amusante pour mon enfant. Ce jeu va motiver mon enfant
Attitude	Je suis favorable à l'utilisation de jeu comme *FunSpeech* pour les enfants. *FunSpeech* est une bonne idée.
Intention d'usage	J'ai l'intention d'utiliser le jeu avec mon enfant après l'étude clinique. Après l'étude clinique, je pense chercher un jeu équivalent à *FunSpeech* pour mon enfant.

Le parent devait déclarer son degré d'accord avec chaque item sur une échelle de Likert en sept points, allant de 1 « *Pas du tout d'accord* » à 7 « *Tout à fait d'accord* », comme présenté ci-dessous.

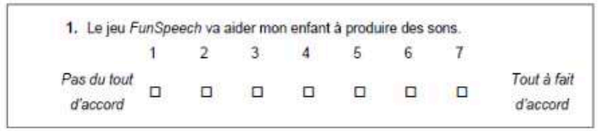


Fig. 6. Exemple d'une affirmation pour l'*Attente de performance*

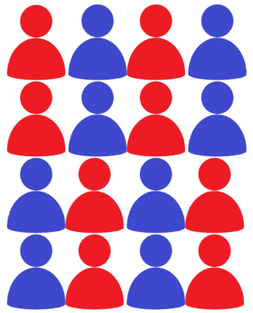
Quand utilité effective et perception de l'utilité diffèrent : le cas d'un jeu sérieux pour l'entraînement post-appareillage d'enfants malentendants

Maëlys Le Magadou[1/2], Bétrancourt Mireille[1], Perez-Fornos Angélica[2], Glück Florent[3], Deriaz Marielle[2], Valentini Gregory[2], Baussand Lou[2], Fritz Matilde[2]

EIAH'25

Contexte : 1 variable indépendante avec 2 niveaux

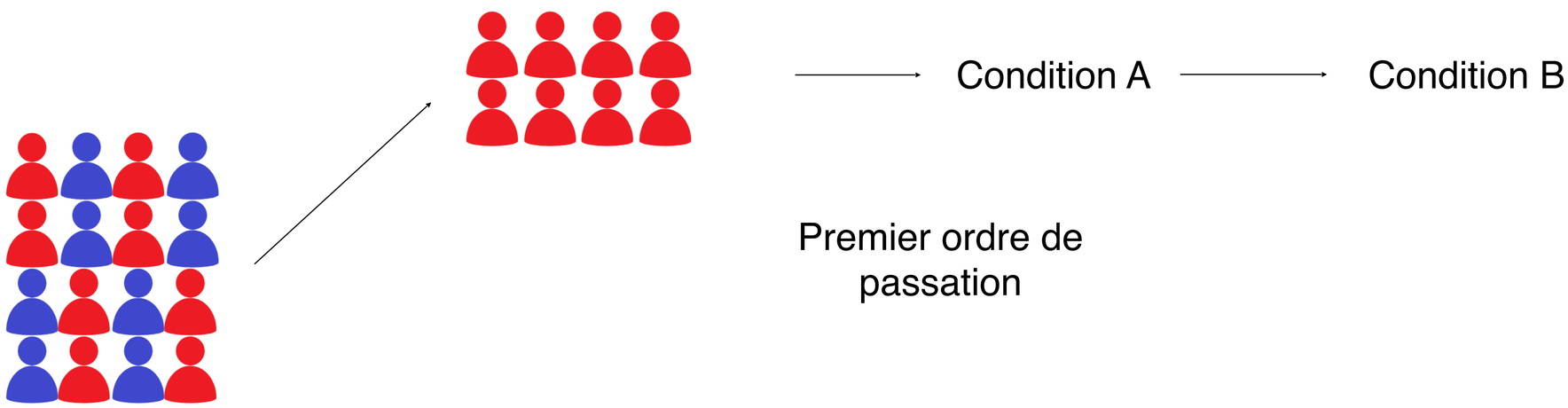
- Protocole intra-sujets



Echantillon initial

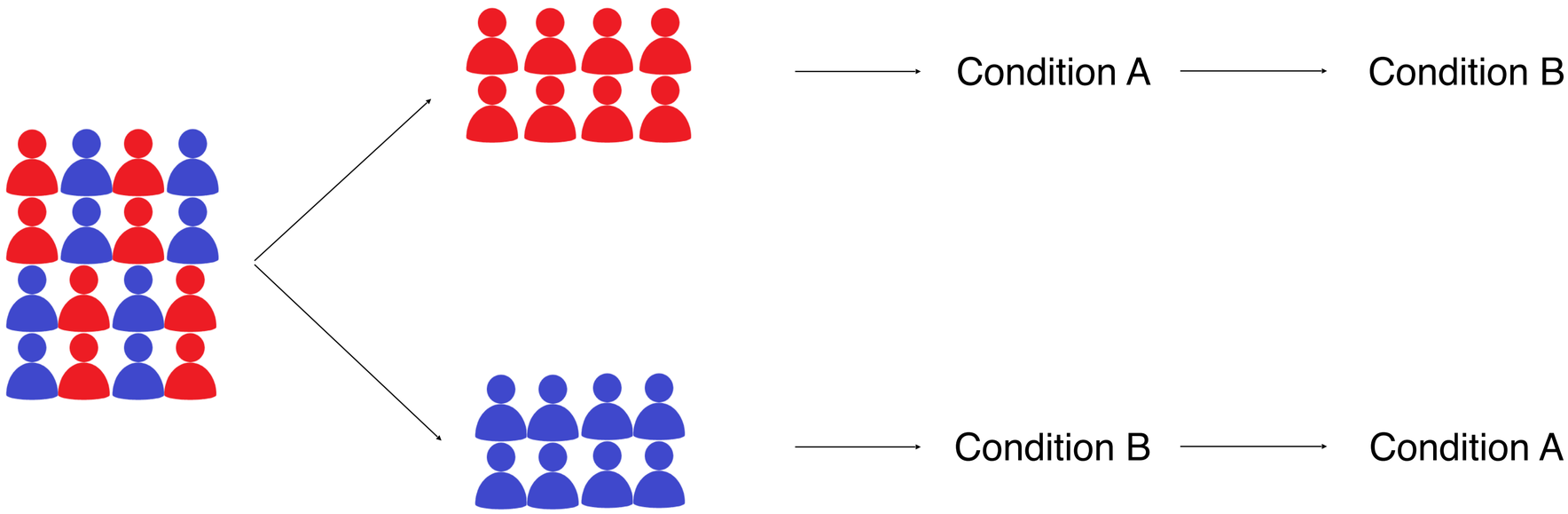
Contexte : 1 variable indépendante avec 2 niveaux

- Protocole intra-sujets



Contexte : 1 variable indépendante avec 2 niveaux

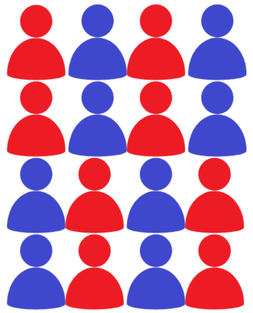
- Protocole intra-sujets



Deuxième ordre de passation

Contexte : 1 variable indépendante avec 2 niveaux

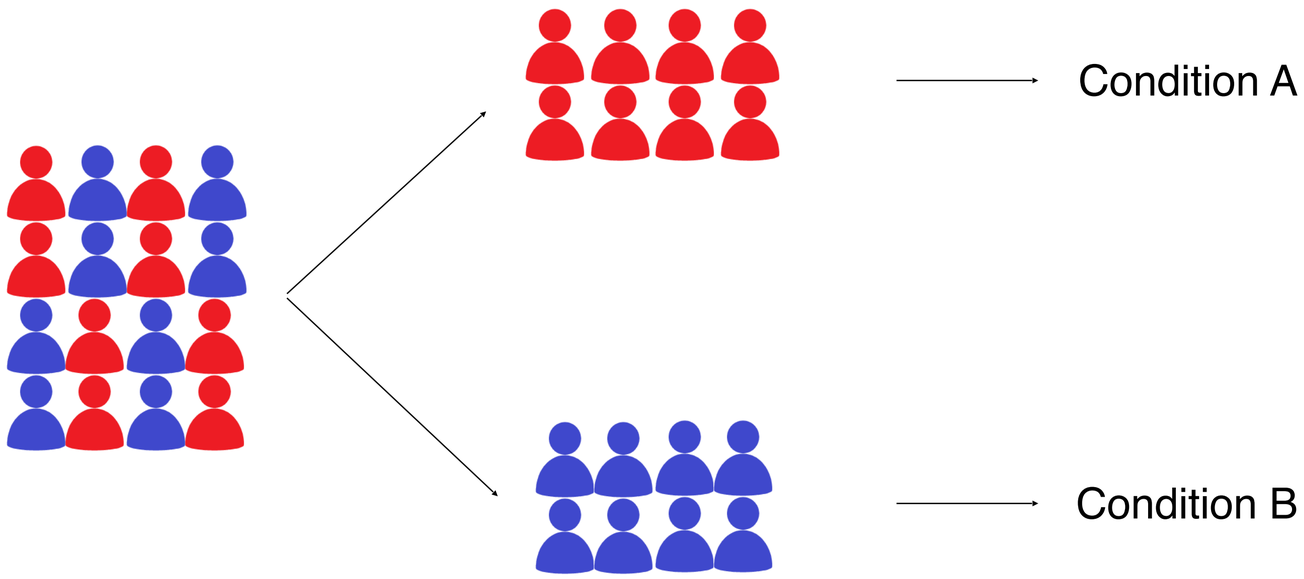
- Protocole inter-sujets



Echantillon initial

Contexte : 1 variable indépendante avec 2 niveaux

- Protocole inter-sujets



Within or Between? Comparing Experimental Designs for Virtual Embodiment Studies

Grégoire Richard[1]* Thomas Pietrzak[1] Ferran Argelaguet[2] Anatole Lécuyer[2] Géry Casiez[1,3]

[1]Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, Lille, France
[2]Inria Rennes - Bretagne Atlantique, Rennes, France [3]Institut Universitaire de France, Paris, France

2022 IEEE Conference Virtual Reality and 3D User Interfaces (VR) | 978-1-6654-9617-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/VR51125.2022.00037

ABSTRACT

When designing virtual embodiment studies, one of the key choices is the nature of the experimental factors, either between-subjects or within-subjects. However, it is well known that each design has advantages and disadvantages in terms of statistical power, sample size requirements and confounding factors. This paper reports a within-subjects experiment with 92 participants comparing self-reported embodiment scores under a visuomotor task with two conditions: synchronous motions and asynchronous motions with a latency of 300 ms. With the gathered data, using a Monte-Carlo method, we created numerous simulations of within- and between-subjects experiments by selecting subsets of the data. In particular, we explored the impact of the number of participants on the replicability of the results from the 92 within-subjects experiment. For the between-subjects simulations, only the first condition for each user was considered to create the simulations. The results showed that while the replicability of the results increased as the number of participants increased for the within-subjects simulations, no matter the number of participants, between-subjects simulations were not able to replicate the initial results. We discuss the potential reasons that could have led to this surprising result and potential methodological practices to mitigate them.

Index Terms: Virtual Embodiment—Methodology—Latency

1 INTRODUCTION

The design of a controlled experiment requires determining the number of independent variables to consider, and the number of levels taken by each of them. For any independent variable involved, the experimenter has to choose between a between-subjects design, where each participant is exposed to only one level, and a within-subjects design where each participant tests each level. A between-subjects design is required when investigating individual differences. For example, the effect of gender or age on a dependent variable, and a within-subjects design is required when studying learning or fatigue effects on an independent variable. However, most of the time the choice of an experimental design is not fixed and it is decided by the researcher.

Each design has its own advantages and disadvantages. Between-subjects designs prevent confounding factors such as learning, fatigue, and frustration since each participant is exposed to a single level. The time taken to complete the experiment is also shorter than a within-subjects design. However, between-subjects designs are based on the comparisons between different groups of participants to study the effect of an independent variable. The results can be affected by individual differences making it harder to detect significant differences, reducing the power of the statistical analysis. It also increases the chances to detect differences that can be attributed to differences between the groups instead of differences between the conditions, when the groups are too small and not homogeneous (*e.g.* typing speed when comparing different typing techniques [50]). To reduce the impact of individual differences, between-subjects designs comparatively require larger groups compared to within-subjects designs (m times larger if m is the number of levels of an independent variable). Mixed experiment designs can be used when considering two or more independent variables, where one or more independent variables are administered within-subjects and the others between-subjects, but it is beyond the scope of this work.

Controlled experiments in virtual reality (VR) to study user experience do not suffer an exception. Some contexts require to follow a within-subjects design, for example when studying the effect of VR therapy over different sessions on a phobia [73]. Other contexts, such as scenarios involving surprises/threats [25, 31], typically follow a between-subjects design as being aware of the effect can alter participants' reactions. Other contexts, such as the study of embodiment, offer more freedom in the choice of an experimental design. A key factor of user experience in VR is virtual embodiment, the "sense that emerges when [a body]'s properties are processed as if they were the properties of one's own biological body" [45]. While a majority of virtual embodiment studies are designed following a within-subjects design, a fair amount follows a between-subjects design (see Table 1). For each design, it remains unclear how many participants should be involved in a study to detect an effect if it exists. Furthermore, to get the same statistical power between the two designs, the number of participants required in a between-subjects design may not be directly related to the number of levels. The behavior of participants when answering successive questionnaires may also influence the results, as they may answer the questions depending on their previous answers, for example.

To answer these questions, we evaluate in this paper the influence of the experimental design and the sample size in a virtual embodiment experiment with a single independent variable with two levels. We first report an experiment ($n = 92$), in which participants had to perform a visuomotor task adapted from previous studies [12,46,59]. The experimental design had one within-subjects variable with two levels, synchronous and asynchronous integration. The task and design were chosen to ensure that a significant difference could be observed between the two levels with lower embodiment scores for the asynchronous condition, as it has been previously shown [12,46,59]. We then evaluated the influence of the number of participants and the type of experimental design on the effect size and the ability to find the expected significant difference between the conditions. We used a Monte-Carlo method on different subsets of the collected data to simulate a number of participants ranging from 10 to 92 in the within-subjects design and 12 to 46 per group in the between-subjects design. Our results show that we could find the expected result with more than 95% chance from 33 participants with the within-subjects design, in agreement with the results from the literature. However, we were not able to replicate the expected result with the between-subjects design. We discuss the implications of these findings and provide guidelines on the design of virtual embodiment studies and embodiment questionnaires.

*e-mail: gregoire.richard@inria.fr

978-1-6654-9617-9/22/$31.00 ©2022 IEEE
DOI 10.1109/VR51125.2022.00037

IEEE VR'22 https://dx.doi.org/10.1109/VR51125.2022.00037

Contexte

- Incarnation virtuelle

Kilteni *et al.*, 2012

  - « Le Sentiment d'Incarnation envers un corps B est le sentiment qui émerge quand les propriétés du corps B sont traitées comme les propriétés de son propre corps biologique. »

Contexte

Reference	Year	Design	Nb levels	Nb participants	Objective measurements	Subjective Measurements		Effect (pt)*
						Number of questions	Reference questionnaire	
[74]	2010	Within	2	14	Proprioceptive drift	11	[11]	1.29
[80]	2010	Between	2x2x2	24	Heart rate	8	-	1.1
[60]	2011	Within	2	22	Behavior	11	[11]	1.25
[66]	2011	Within	2	20	GSR	9	-	0.75
[7]	2013	Between	2	30	Performance	4	-	1.25
[59]	2013	Between	2x2	18	Heart rate	7	[53, 80]	1
[46]	2014	Between	2x2	60	GSR	8	[87]	1.25
[61]	2014	Within	2x2	15	Proprioceptive drift	12	[11, 74]	0.38
[89]	2015	Between	3	53	Performance	10	-	0.21
[58]	2015	Within	2x2x2	43	Proprioceptive drift/GSR	13	[11, 43, 79]	2
[41]	2015	Within	6	13	Proprioceptive drift	2	-	1.4
[2]	2016	Within	3	33	Behavior and movement	8	-	0.5
[8]	2016	Between	2x3	60	Behavior	5	-	0.75
[40]	2016	Between	2	24	Performance	12	[52, 80]	0.3
[85]	2016	Within	5	10	Performance	2	[41]	-
[23]	2017	Mixed	2x2	48	GSR	10	-	0.75
[12]	2017	Between	2	32	Performance/Behavior	7	[52, 80]	1
[67]	2017	Within	2x2x2	39	Behavior	14	[55]	0.9
[21]	2017	Within	3x3x2x1	23	-	6	[2, 57, 80]	1.4
[37]	2017	Within	2	28	Performance	10	[2, 24]	0.75
[75]	2017	Within	2x4	24	-	32	[87]	-
[5]	2017	Within	2x2	22	Behavior	12	[55]	2.5
[30]	2018	Within	3	20	Performance	9	[43]	1.5
[1]	2018	Within	4	25	-	18	[2]	3.75
[32]	2018	Within	3	32	Proprioceptive drift	10	[55]	*0.8***
[42]	2018	Within	2	10	Behavior	16	[6]	1
[56]	2018	Between	3	75	Performance	13	[72]	0.35
[77]	2018	Within	5x2	41	Performance	14	[11, 58]	1.15
[13]	2018	Within	3x2	21	-	2	-	-
[48]	2019	Between	5	60	GSR	4	[79]	2
[38]	2019	Within	3x2	34	Performance	12	[2, 24, 37]	0.25
[76]	2019	Within	3x3	37	Performance	5	[77]	0.55
[81]	2019	Within	2	16	-	16	[43]	0.45
[14]	2019	Within	2	30	Performance	12	-	1.3
[4]	2019	Between	2x2	64	Behavioral	4	[9]	1.1
[31]	2020	Mixed	2x3	30	Behavior	14	[35]	1.5
[65]	2020	Between	2x2	60	Performance	11	[35]	1.5
[64]	2021	Within	2x2x2	99	Performance	24	[35]	-
[68]	2021	Within	3	24	Performance	14	[35]	0.35
[78]	2021	Mixed	2x2	74	Performance	4	[17]	0.6

Objectifs de l’étude

- **Comparer les principaux protocoles expérimentaux dans un contexte d’incarnation virtuelle**
  - Les protocoles inter- et intra-sujets ont-ils la même capacité à détecter des effets ?
  - Dans quelle mesure la taille de l’échantillon influence-t-elle cette capacité ?

Protocole expérimental

- Réplication de résultats de la littérature portant sur l’incarnation virtuelle

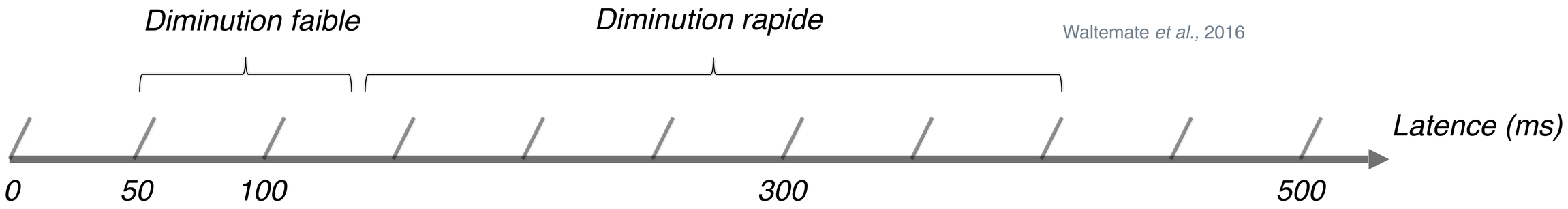
- 2 conditions expérimentales jouant sur l’impact de la latence sur l’incarnation

Contexte

- Incarnation virtuelle

Kilteni *et al.*, 2012

  - « Le Sentiment d'Incarnation envers un corps B est le sentiment qui émerge quand les propriétés du corps B sont traitées comme les propriétés de son propre corps biologique. »



Protocole expérimental

Protocole expérimental

92 participants

Configuration expérimentale, avec casque de RV, manettes et trackers, équipé par un participant (gauche), les deux avatars proposes (milieu), et l'environnement virtuel avec l'avatar masculin (droite).

Protocole expérimental

	1ère condition	2ème condition
Condition synchrone	*G1*	*G2*
Condition asynchrone	*G2*	*G1*

G1 : 46 participants

G2 : 46 participants

Total : 92 participants

ID	Questions
R1	I felt out of my body.
R2	I felt as if my (real) body was drifting toward the virtual body or as if the virtual body was drifting toward my (real) body.
R3	I felt as if the movements of the virtual body were influencing my own movements.
R4	It felt as if my (real) body was turning into an "avatar" body.
R5	At some point it felt as if my real body was starting to take on the posture or shape of the virtual body that I saw.
R6	I felt like I was wearing different clothes from when I came to the laboratory.
R7	I felt as if my body had changed.
R9	I felt that my own body could be affected by the virtual body.
R10	I felt as if the virtual body was my body.
R11	At some point, it felt that the virtual body resembled my own (real) body, in terms of shape, skin tone or other visual features.
R12	I felt as if my body was located where I saw the virtual body.
R13	I felt I could control the virtual body as if it was my own body.
R14	It seemed as if I felt the touch of the table in the location where I saw the virtual legs touched.
R15	It seemed as if the touch I felt was caused by the table touching the virtual body.
R16	It seemed as if my body was touching the table.

Protocole expérimental

	1ère condition	2ème condition
Condition synchrone	*G1*	*G2*
Condition asynchrone	*G2*	*G1*

intra-sujets

Protocole expérimental

	1ère condition	2ème condition
Condition synchrone	*G1*	*G2*
Condition asynchrone	*G2*	*G1*

inter-sujets

Résultats globaux sur l'incarnation

- Synchrone > Asynchrone pour le protocole intra-sujets

- Pas de différence significative pour le protocole inter-sujets (p = 0.23)

Metric	p-value	V	Sync.		Async.	
			Mdn	IQ	Mdn	IQ
Response	< 0.001*	743	0.0	$[-0.8, 0.7]$	-0.2	$[-1.4, 0.6]$
Ownership	< 0.001*	623.5	0.83	$[0.25, 1.42]$	0.16	$[-0.5, 0.92]$
Appearance	$= 0.07$	1528.5	-0.38	$[-1.13, 0.5]$	-0.69	$[-1.25, 0.38]$
Multi-sensory	< 0.001*	922	1.33	$[1.0, 1.75]$	1.17	$[0.34, 1.5]$
Embodiment	< 0.001*	959.5	0.42	$[0.01, 1.00]$	0.08	$[-0.56, 0.74]$

Influence de la taille de l’échantillon

- Simulation d’expérience virtuelle avec une méthode de Monte-Carlo (approximation par grand nombre de tirages)

Influence de la taille de l’échantillon

- Simulation d’expérience virtuelle avec une méthode de Monte-Carlo (approximation par grand nombre de tirages)
    - Echantillon de 10 à 92 participants

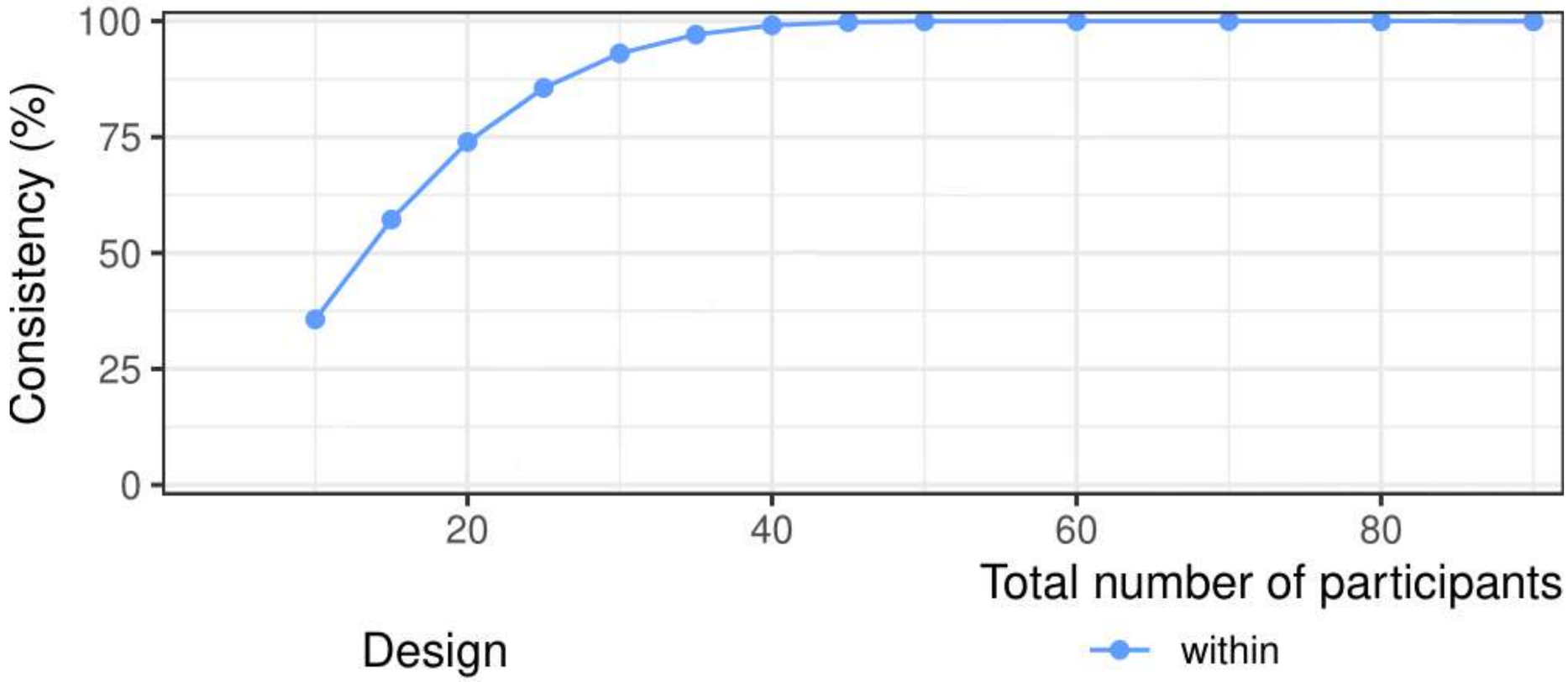
Influence de la taille de l’échantillon

- Simulation d’expérience virtuelle avec une méthode de Monte-Carlo (approximation par grand nombre de tirages)
  - Echantillon de 10 à 92 participants
  - Conformité de chaque expérience virtuelle

Influence de la taille de l’échantillon

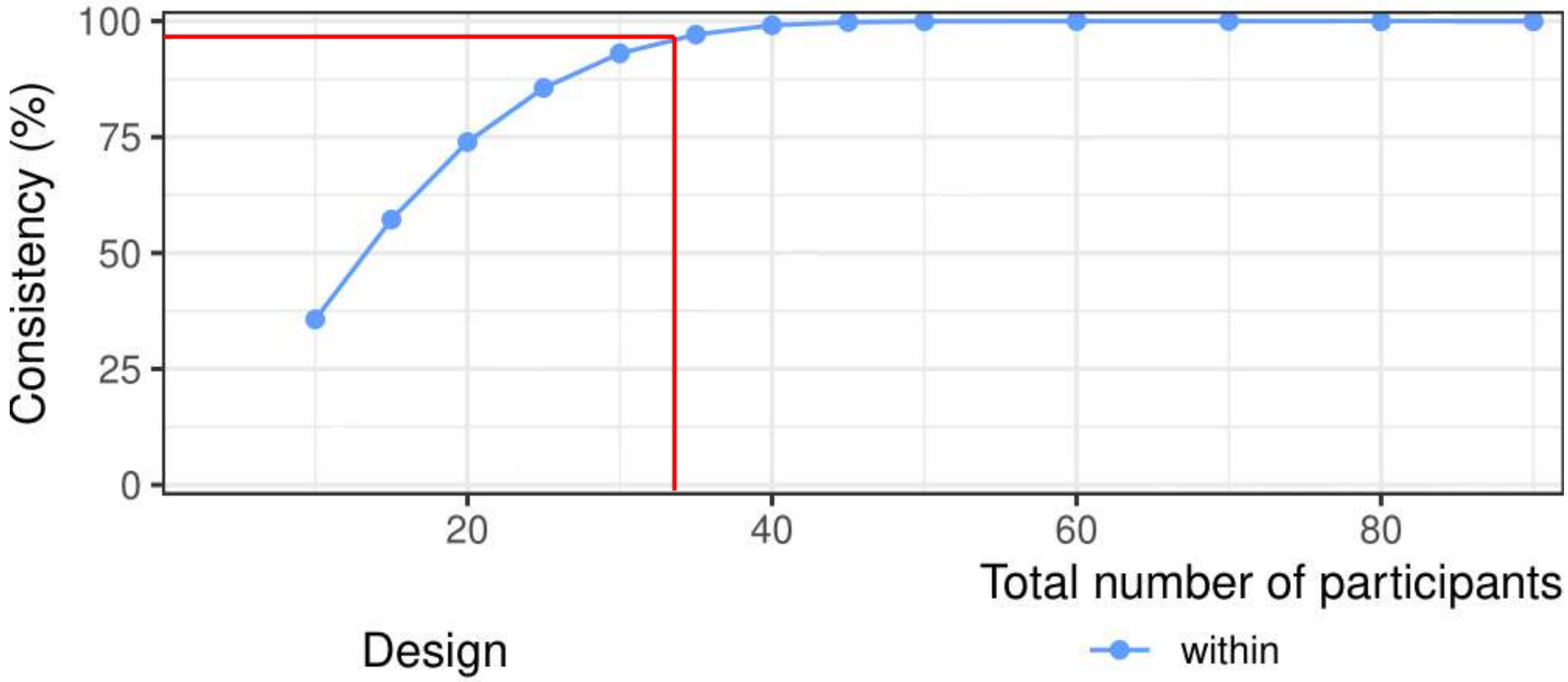
- Simulation d’expérience virtuelle avec une méthode de Monte-Carlo (approximation par grand nombre de tirages)
  - Echantillon de 10 à 92 participants
  - Conformité de chaque expérience virtuelle
  - Taille de l’effet

Influence de la taille de l’échantillon



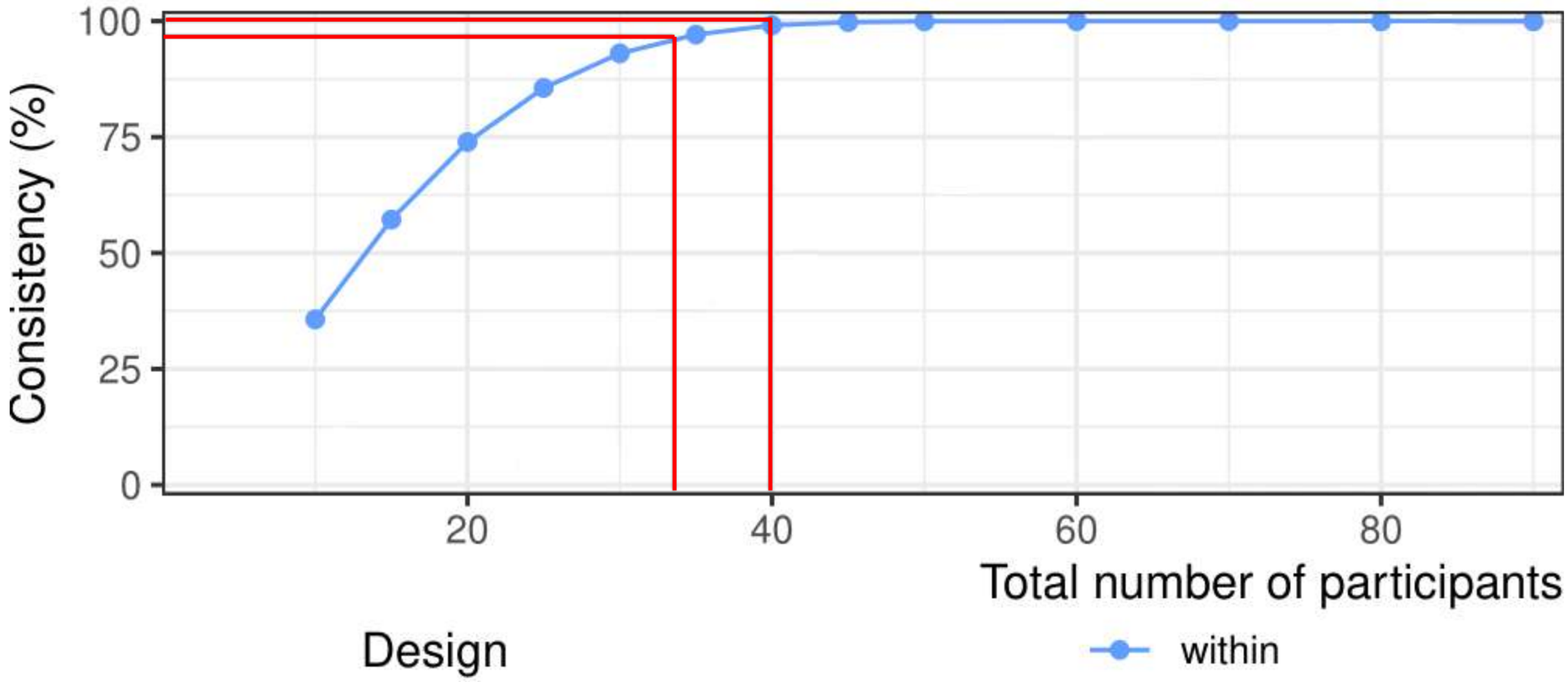
Conformité de résultats en fonction de la taille de l’échantillon pour les experiences virtuelles en protocole intra-sujets.

Influence de la taille de l’échantillon



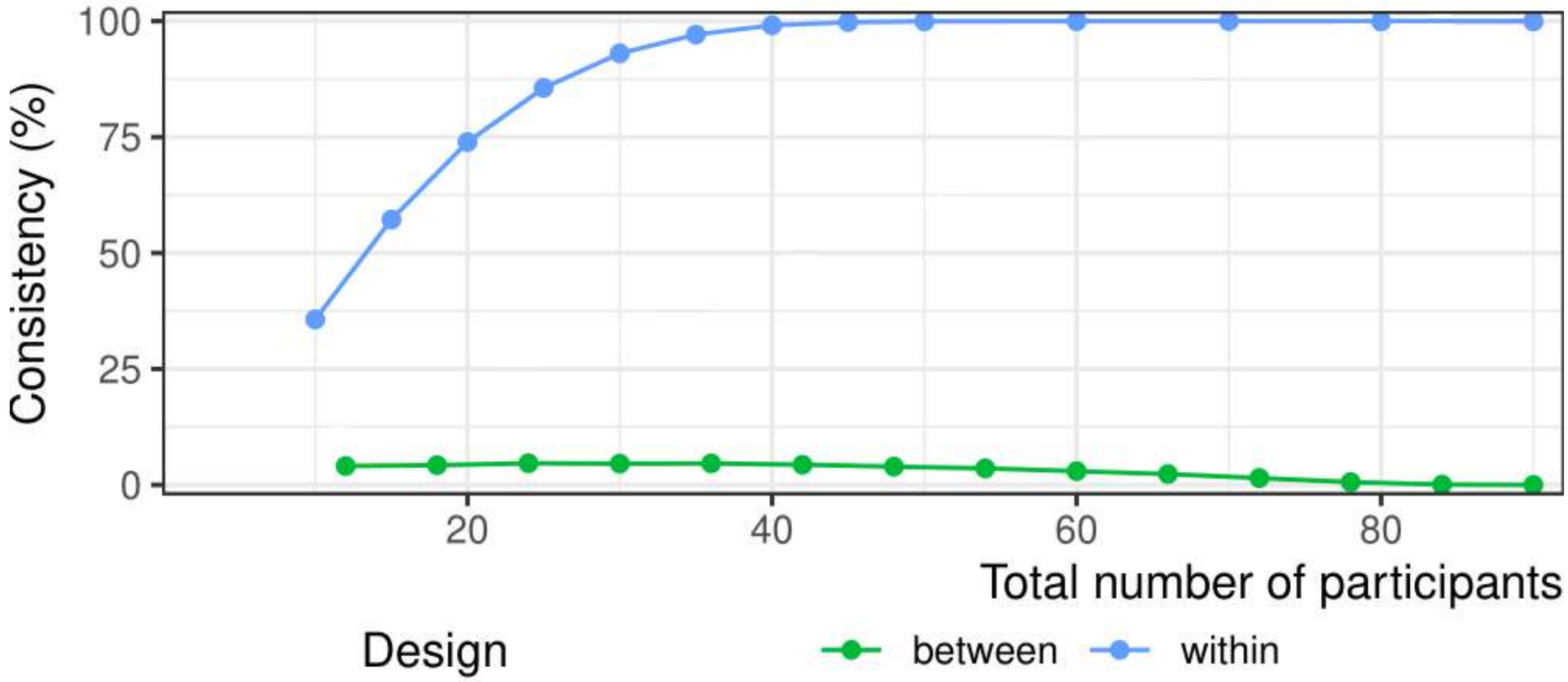
Conformité de résultats en fonction de la taille de l’échantillon pour les experiences virtuelles en protocole intra-sujets.

Influence de la taille de l’échantillon



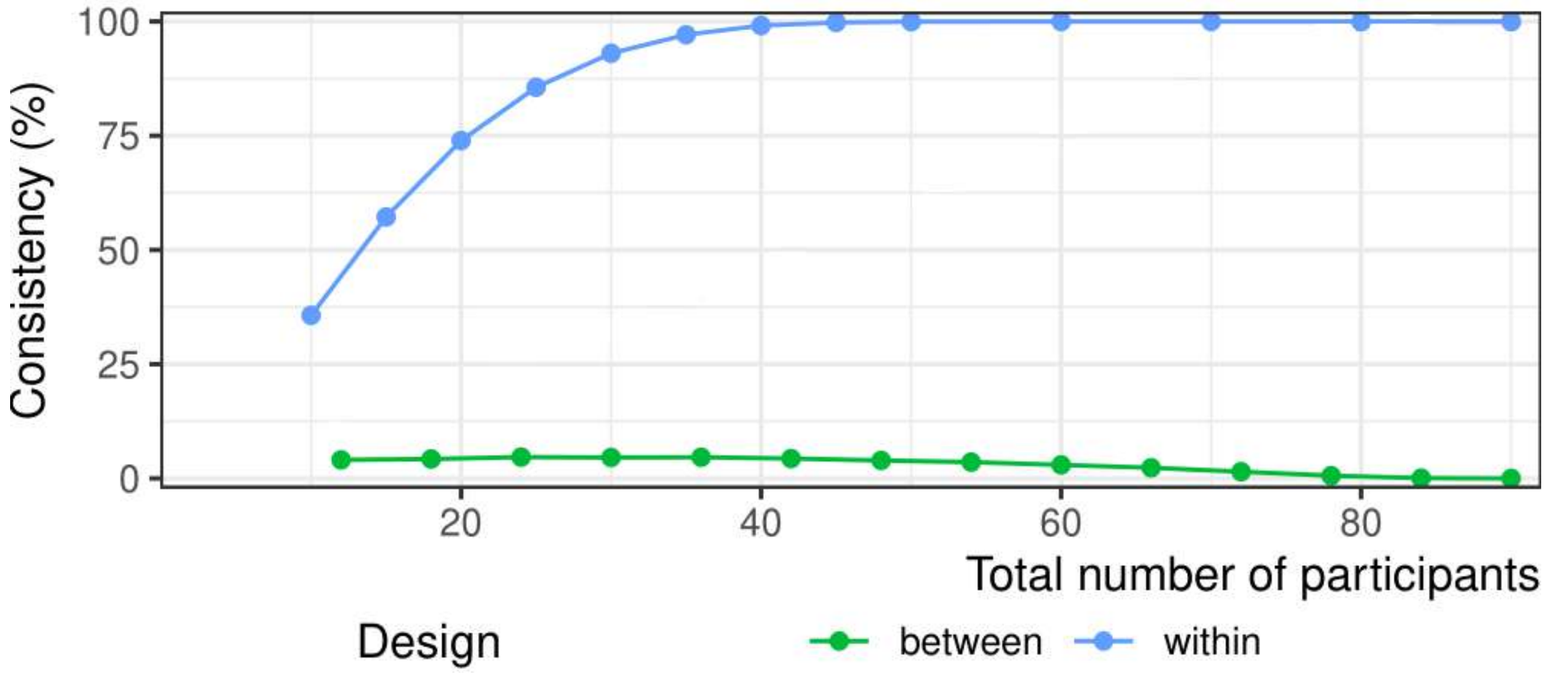
Conformité de résultats en fonction de la taille de l’échantillon pour les experiences virtuelles en protocole intra-sujets.

Influence de la taille de l’échantillon



Conformité de résultats en fonction de la taille de l’échantillon pour les experiences virtuelles en protocole intra-sujets (bleu), et en protocole inter-sujets (vert).

Influence de la taille de l’échantillon

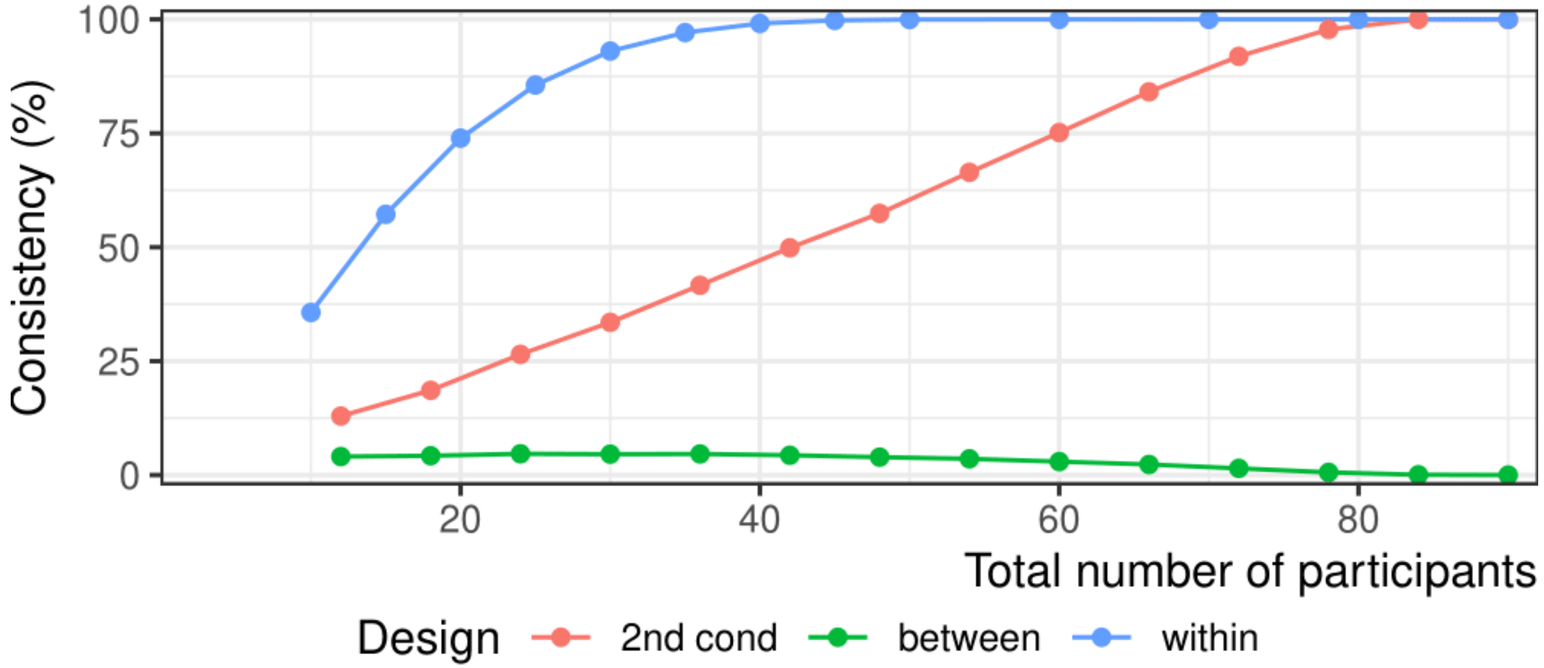


Conformité de résultats en fonction de la taille de l’échantillon pour les experiences virtuelles en protocole intra-sujets (bleu), et en protocole inter-sujets (vert).

	1ère condition	2ème condition
Condition synchrone	*G1*	*G2*
Condition asynchrone	*G2*	*G1*

inter-sujets
2ème condition

Influence de la taille de l'échantillon

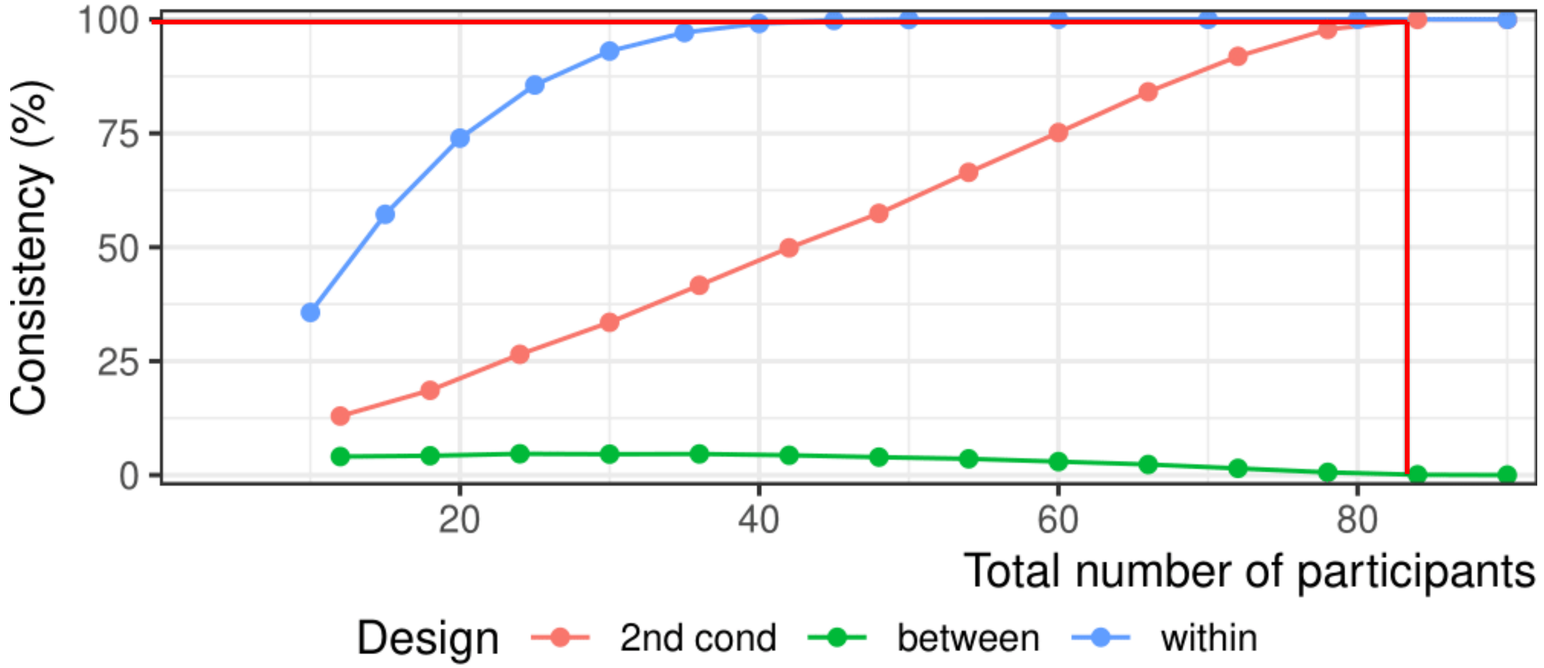


Conformité de résultats en fonction de la taille de l'échantillon pour les experiences virtuelles en protocole intra-sujets (bleu), et en protocole inter-sujets (vert), et l'inter-sujets 2ème condition (rouge).

	1ère condition	2ème condition
Condition synchrone	*G1*	*G2*
Condition asynchrone	*G2*	*G1*

inter-sujets 2ème condition

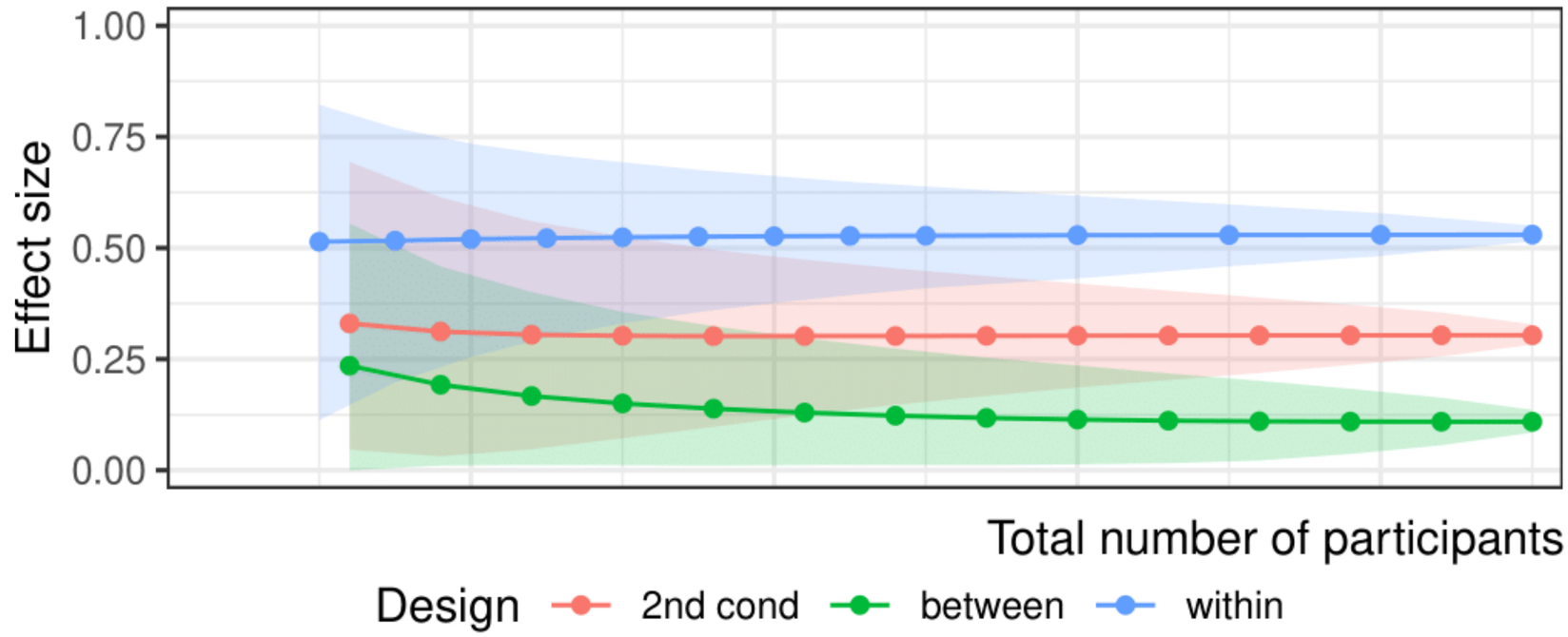
Influence de la taille de l’échantillon



	1ère condition	2ème condition
Condition synchrone	*G1*	*G2*
Condition asynchrone	*G2*	*G1*
		inter-sujets 2ème condition

Conformité de résultats en fonction de la taille de l’échantillon pour les experiences virtuelles en protocole intra-sujets (bleu), et en protocole inter-sujets (vert), et l’inter-sujets 2ème condition (rouge).

Influence de la taille l’échantillon sur la taille d’effet



Taille d’effet moyen en fonction du nombre de participants pour le protocole intra- (bleu), inter- (vert), et inter-sujets 2ème condition (rouge).

Discussion

effectsize measure	between groups			within groups		
	small	medium	large	small	medium	large
r, ω, η	0.14	0.27	0.61	0.12	0.32	0.72
η^2, R^2, ω^2, ϵ^2	0.02	0.07	0.37	0.01	0.10	0.52
Cohen's d	0.22	0.44	1.09	0.19	0.52	1.38
OR	1.48	2.13	5.90	1.39	2.44	8.72
CLES	0.54	0.59	0.71	0.54	0.60	0.76
Cohen's f	0.11	0.22	0.55	0.10	0.26	0.69
Cramer's V	0.10	0.19	0.42	0.08	0.22	0.50
f^2	0.02	0.08	0.59	0.01	0.11	1.08
Kendall's τ	0.09	0.17	0.41	0.08	0.20	0.51
Spearman's ρ	0.13	0.26	0.58	0.11	0.30	0.69

Small, Medium, Large? A Meta-Study of Effect Sizes at CHI to Aid Interpretation of Effect Sizes and Power Calculation [CHI'25]

- Comment expliquer la non-réplication des résultats pour le protocole inter-sujets ?
  - Influence de la taille de l'effet sur les résultats
  - Nombre de participants en inter-sujets
  - Changement de comportements et de stratégie pour répondre aux questionnaires d'incarnation
  - Mesure absolue de l'incarnation avec les questionnaires ?

Implications pour l’administration de questionnaires

- Présenter le questionnaire en début d’expérience ?
- Présenter des conditions permettant d’expérimenter les extrêmes de chaque question ?
- Présentation des réponses du questionnaire précédent ?
- Changer l’échelle des questionnaires ?
- Besoin de mieux comprendre la stratégie suivie par les participants pour répondre aux questionnaires
- Nécessité de mieux détailler comment les questionnaires ont été administrés

Conclusion

- Préférer les protocoles intra-sujets aux inter-sujets
- Nécessité de développer de nouveaux outils méthodologiques pour les protocoles inter-sujets
- Etudier si le phénomène existe avec d’autres questionnaires