EURO ISME
International Society for Military Ethics in Europe
registered in FRANCE as ASSOCIATION loi 1901 - W563003729

La perception de l'agentivité et les valeurs morales relatives aux armes autonomes:

Étude empirique utilisant l'approche de la conception éthique

par:
Ilse Verdiesen

Mémoire EuroISME de l'année 2018

La perception de l'agentivité et les valeurs morales relatives aux armes autonomes:

Étude empirique utilisant l'approche de la conception éthique

Commandant Ilse Verdiesen

Traduction française: Col. Jean-François Palard

Image de couverture:

Une partie de ce mémoire est basée sur les articles suivants: Verdiesen, I., Dignum, V., & Rahwan, I. (2018, Septembre). Design Requirements for a Moral Machine for Autonomous Weapons. In International Conference on Computer Safety, Reliability, and Security (pp. 494-506). Springer, Cham.
https://link.springer.com/chapter/10.1007/978-3-319-99229-7_44

Verdiesen, I., de Sio, F. S., & Dignum, V. (2019). Moral Values Related to Autonomous Weapon Systems: An Empirical Survey that Reveals Common Ground for the Ethical Debate. IEEE Technology and Society Magazine, 38(4), 34-44.
https://ieeexplore.ieee.org/abstract/document/8924586

La perception de l'agentivité et les valeurs morales relatives aux armes autonomes:

Étude empirique utilisant l'approche de la conception éthique

par:
Commandant Ilse Verdiesen

Étudiante à:

Université de Technologie de Delft,
Les Pays Bas

Le jury pour l'attribution du prix annuel du meilleur mémoire en éthique militaire comprend des membre suivants:

1. Colonel (ret) le Révérend. Professeur Dr. P.J. McCormack, Membre de l'Ordre de l'Empire britannique, (Président, Royaume-Uni)

2. Lieutenant-Colonel (retraité) Dr. Daniel Beaudoin (France/Israel)

3. Dr. Veronika Bock (Allemagne)

4. Aumônier en chef Stefan Gugerel (MA; Autriche)

5. Colonel (retraité) Professeur Dr. Boris N. Kashnikov (Russie)

6. Dr. Asta Maskaliūnaitė (Estonie)

7. Professeur Dr. Desiree Verweij (Pays Bays)

- Mme. Ivana Gošić (Serbie, secretaire)

Renseignements: secretariat.ethicsprize@euroisme.eu
www.euroisme.eu

Le prix à reçu le soutien de:

Sommaire

Les armes autonomes sont de plus en plus utilisées sur les champs de bataille (Roff, 2016). Les systèmes autonomes peuvent présenter de nombreux avantages dans le domaine militaire, par exemple lorsque le pilote automatique d'un F-16 empêche un crash (NOS, 2016), ou dans l'emploi de robots par les démineurs pour neutraliser des bombes (Carpenter, 2016). Cependant la nature même des armes autonomes peut aussi conduire à des actions incontrôlables et à des troubles d'ordre sociétal. L'emploi d'armes autonomes sur les champs de bataille sans contrôle humain direct est non seulement une révolution dans le domaine militaire selon l'opinion de Kaag et Kaufman (2009), mais peut également être considéré comme une révolution morale. Comme l'emploi massif de l'intelligence artificielle sur les champs de bataille semble inévitable (Rosenberg & Markoff, 2016), une recherche portant sur la responsabilité éthique et morale est impérative.

Dans le débat sur les Armes Autonomes des opinions et des points de vue se font fortement entendre. La Campagne pour la Neutralisation des robots tueurs (2017) affirme par exemple sur son site internet: «Le fait de laisser des décisions sur la vie et la mort être prises par des machines transgresse une limite morale fondamentale. Les robots autonomes sont dépourvus de jugement humain et de la capacité à comprendre un contexte». Nous avons trouvé peu de travaux de recherche empiriques qui vont dans le même sens, ou qui donnent un aperçu de la manière dont les Armes Autonomes sont perçues par le grand public et par les militaires. Nous n'avons pas trouvé non plus de recherches empiriques sur les valeurs morales qui sous-tendent la «limite morale fondamentale» des Armes Autonomes. Par conséquent, le vide de connaissances

est double par le fait que l'on n'a pas d'aperçu sur (1) la manière dont les Armes Autonomes sont perçues par les militaires et le grand public et (2) le fait de savoir quelles sont les valeurs morales que l'on trouve importantes si les Armes autonomes sont employées dans un avenir proche.

La première partie de ce manque de connaissances peut être comblée en étudiant la perception des Armes Autonomes en utilisant la théorie de l'agentivité [1] proposée dans les domaines de la psychologie cognitive, de l'intelligence artificielle et de la philosophie morale. La seconde partie du manque de connaissances peut être comblée en étudiant les théories des valeurs connues (Beauchamp & Walters, 1999; Friedman & Kahn Jr, 2003; Schwartz, 2012) pour déterminer quelles sont les valeurs jugées importantes dans l'emploi des armes autonomes. Partant de ce manque de connaissances et de l'énoncé de ces problèmes, nous avons élaboré le sujet de recherche suivant pour cette étude:

Comment les armes autonomes sont-elles perçues par le grand public et les personnels militaires du Ministère de la Défense néerlandais, et quelles valeurs morales considèrent-ils importantes dans l'emploi des armes autonomes?

Dans cette étude, nous appliquerons la méthode basée sur la conception éthique («value-sensitive design», VDS) comme approche de recherche. La VDS est une approche à trois volets qui tient compte des valeurs humaines dans tout le processus de la conception technologique. Il s'agit d'un processus itératif

[1] (Note du traducteur) «agency» en anglais. En **philosophie**, l'**agency** (terme récemment traduit par *agentivité*, notamment au Canada) est la faculté d'**action** d'un **être**; sa capacité à agir sur le monde, les choses, les êtres, à les transformer ou les influencer. En **sociologie**, l'agentivité est la capacité d'agir, par opposition à ce qu'impose la **structure**.

pour la recherche *conceptuelle, empirique et technologique* des valeurs humaines impliquées dans la conception (Davis & Nathan, 2015; Friedman & Kahn Jr, 2003). La recherche conceptuelle s'articule en deux parties: (1) l'identification des parties prenantes *directes*, les futurs utilisateurs de la technologie, et les parties prenantes *indirectes*, dont la vie est affectée par la technologie, et (2) l'identification et la définition des valeurs que l'emploi de la technologie concernera. La recherche empirique s'intéresse à la compréhension et au vécu des parties prenantes dans les contextes ayant rapport avec la technologie, et les valeurs concernées seront examinées. Dans la recherche d'ordre technique, les caractéristiques spécifiques de la technologie sont analysées (Davis & Nathan, 2015).

Nous nous éloignons légèrement de la méthode de la conception éthique («value-sensitive design», VDS), car nous ne procédons pas à une analyse complète des parties prenantes pour les identifier dans le volet 1, mais nous nous concentrons sur deux groupes proéminents, le grand public et les personnels militaires, car ces deux groupes étaient réactifs pour notre étude. Pour le volet 2 de la première phase, nous avons procédé à un examen des publications afin d'identifier les valeurs concernées par les armes autonomes. L'objectif principal de notre recherche est la phase d'investigation, car nous avons trouvé que le volet empirique est passé sous silence dans le débat éthique sur les armes autonomes. Dans la phase d'investigation technique, nous ne désignons pas une arme autonome de la manière dont on pourrait intuitivement s'y attendre, car cela constituerait un projet immense qui dépasserait le propos de cette étude. Cependant nous avons décidé d'utiliser le travail du groupe de recherche de coopération évolutive et de proposer la conception d'une «machine morale» pour armes autonomes qui peut être utilisée

dans une étape ultérieure de la recherche sur l'éthique des armes autonomes.

La pertinence scientifique de notre étude est que nous contribuons aux publications académiques pour mieux comprendre la perception du grand public et des militaires vis-à-vis des armes autonomes, et pour identifier les valeurs morales que le grand public et les militaires associent aux armes autonomes. Actuellement cette compréhension est insuffisante, et aucune donnée empirique sur la perception et les valeurs associées aux armes autonomes n'a pu être trouvée. En utilisant l'approche de recherche de la conception éthique nous démontrons que cette méthode est applicable pour structurer une recherche académique pouvant être considérée comme étude de cas pour l'approche VSD. Nous approfondissons également l'étude de la prise de décision éthique concernant les vecteurs autonomes effectuée par Bonnefon, Shariff et Rahwan (2016) pour l'appliquer au domaine des armes autonomes.

La pertinence d'un point de vue sociétal est que la compréhension de la perception qu'ont le grand public et les personnels militaires du Ministère de la Défense néerlandais des armes autonomes, et la détermination des valeurs qu'ils associent à l'emploi des armes autonomes, peuvent être utilisées pour trouver des points communs et des différences dans le débat portant sur cette technologie lancé par la Campagne pour arrêter les Robots Tueurs (2015) et la Commission Internationale pour le Contrôle des Armes Robotisées (ICRAC). En deuxième lieu, les résultats de cette étude montrent comment la méthode de conception éthique peut être appliquée aux armes autonomes pour identifier les valeurs que le grand public et les personnels militaires associent à l'emploi de ce type d'armes. Enfin, en identifiant les valeurs importantes à incorporer dans la conception des armes

autonomes, l'étude contribue à la conception future et à l'emploi responsable des armes autonomes.

Notre étude se concentre sur la phase d'investigation empirique de la conception éthique. Cette phase s'articule en deux parties. La première partie est exploratoire; c'est l'identification des valeurs des armes autonomes par une recherche en ligne, complétée par des entretiens avec six experts. La deuxième partie de l'investigation empirique étudie la perception de l'agentivité par le biais d'une méthode de recherche confirmatoire. Nous faisons fonctionner le concept d'agentivité pour confirmer notre hypothèse de la perception de l'agentivité en effectuant des expériences contrôlées randomisées, et utilisons les résultats de l'expérience pour explorer les valeurs associées aux armes autonomes de manière descriptive.

Pour étudier la perception de l'agentivité des armes autonomes nous avons émis une hypothèse dans laquelle nous soutenons que les personnels militaires considéreront les armes autonomes comme n'importe quelle autre arme, et donc comme rien de plus qu'un instrument pour obtenir un effet. Nous assumons dans l'étude finale que les personnels militaires considéreront les armes autonomes comme ne possédant aucun état mental. Nous avons élaboré trois scénarios dans nos expériences décrivant (1) un drone actionné par l'homme, (2) une arme autonome n'ayant aucune caractéristique d'agentivité, et (3) une arme autonome ayant des caractéristiques d'agentivité. Nous ne supposions aucune différence entre le statut d'arme autonome à agentivité neutre et le statut d'un drone activé à distance par un être humain. Nous supposons aussi que le statut d'agentivité neutre sera jugé très différent du statut de haute agentivité, dans lequel nous disons précisément aux participants que l'arme autonome a des caractéristiques liées à l'agentivité, comme la capacité de planifier et de fixer ses

propres objectifs. En se basant sur cette démarche, nous supposons que:

H1: les personnels militaires ne considéreront pas que les armes autonomes ont des états mentaux.

Les résultats de l'étude pilote et des études finales montrent que les éléments d'agentivité sont fiables et forment un tout. L'ensemble agentivité comprend les quatre éléments *Pensée, Détermination des objectifs, Libre arbitre* et *Réalisation des objectifs* pouvant être utilisés pour mesurer la perception d'agentivité des armes autonomes et des drones pilotés à distance par l'homme. Notre analyse centrale porte sur la manière dont le scénario de l'agentivité neutre diffère des scénarios Actionné par l'Homme et Haute Agentivité des armes autonomes. Nos résultats indiquent que la perception de l'agentivité par les personnels militaires et les civils travaillant pour le Ministère de la Défense néerlandais vis-à-vis du scénario d'agentivité neutre est plus haute que la perception de l'agentivité vis-à-vis du scénario du drone actionné par l'homme. Cela signifie qu'ils attribuent plus d'agentivité à une arme autonome qu'à un drone actionné par l'homme. A partir de ces conclusions, nous devons rejeter notre hypothèse.

L'effet de la perception de l'agentivité sur les variables dépendantes est étudié de manière descriptive, et non analysé de manière régressive. Nous avons trouvé que 7 sur 9 variables dépendantes sont associées de manière significative à l'agentivité. Il s'agit des variables suivantes: *confiance, dignité humaine, assurance, attentes, soutien (approbation), équité* et *inquiétude.* Nos conclusions sont les suivantes:

- Les personnels militaires et les civils travaillant pour la Défense néerlandaise ressentent plus de confiance, d'assurance et d'approbation (soutien) pour des actions

accomplies par des drones à contrôle humain que pour celles accomplies par des armes autonomes;

- Un drone actionné par l'homme est considéré comme respectant davantage la *dignité humaine* qu'une arme à agentivité neutre ou élevée, même si les actions et le résultat des scénarios sont identiques;
- Les personnels militaires et les civils travaillant pour la Défense néerlandaise ont un même niveau d'attentes vis-à-vis des actions futures des drones actionnés par l'homme que vis-à-vis des armes à agentivité neutre ou élevée. Ils considèrent aussi les actions des drones actionnés par l'homme et celles des armes autonomes comme également équitables;
- Les armes autonomes provoquent plus d'inquiétudes chez les personnels militaires et les civils travaillant pour la Défense néerlandaise que les armes actionnées par l'homme.

Cette étude a été menée à partir d'un ensemble de données réduit et, pour être en mesure de donner aux résultats une portée générale, il faudrait plus de personnes interrogées, et représentant un groupe démographique plus large. Par conséquent, nous proposons que le projet de «Machine morale pour les armes autonomes» prenne la forme d'une «expérience en ligne de grande ampleur» pour être plus exhaustif sur cette question. Pour ce projet nous travaillons sur le concept de la «machine morale» qui a été développé par le Groupe de Coopération Evolutive au Media Lab du MIT (Massachusetts Institute of Technology) (2016). Du fait de la sensibilité de ce sujet, nous proposons une mise en œuvre en deux étapes: d'abord concevoir une expérience contrôlée avec un ensemble limité de conditions (statuts). Ces conditions peuvent être visualisées dans des scénarios qui permettent aux utilisateurs d'y accéder après l'obtention d'un mot de passe par

l'intermédiaire d'une interface web pour un serveur sécurisé. Après avoir rassemblé les données initiales et les réactions des utilisateurs, l'étape suivante pourrait être d'accéder à une plateforme ouverte à plus grande échelle, comme la «machine morale», où les personnes évalueraient les scénarios pour collecter un grand nombre de données de différents groupes démographiques; celles-ci pourraient être utilisées pour obtenir des résultats plus solides ayant une portée plus générale.

On peut identifier plusieurs problèmes posant des limites à cette étude. Tout d'abord, l'opérationnalisation des éléments et le concept d'agentivité dérivent d'une catégorie de publications qui décrivent les caractéristiques de l'agentivité, et notre sélection des caractéristiques a été basée sur un relevé numérique, et non guidée par la pertinence ou des critères de pondération. La seconde limite est la sélection des valeurs à partir du questionnaire sur les valeurs et des entretiens d'experts comme variables dépendantes. Bien que cette sélection fût longuement discutée entre les trois chercheurs concernés dans cette étude, le choix final a été effectué dans le cadre d'une méthode heuristique et non objective. En troisième lieu, les échantillons utilisés dans cette étude sont limités de par leur taille et leur portée démographique; l'étude finale n'a concerné que 239 personnes interrogées, ce qui ne représente qu'une petite partie du personnel de la Défense. Enfin, la répartition des personnes interrogées pour le scénario final est faussée, et le second scénario (armes autonomes à agentivité neutre) concerne 32 personnes interrogées de plus que pour le scénario un (drone actionné par l'homme). L'une des causes de cette dissymétrie pourrait être le fait que les personnes désiraient s'informer sur les armes autonomes, mais renoncèrent lorsqu'on leur présenta le scénario «action humaine». C'est ce que l'on appelle «l'attrition sélective»; elle a un impact négatif sur la validité interne de l'étude. Une autre

explication pourrait être que le logiciel était défaillant en répartissant les personnes interrogées sur les scénarios, ce qui voudrait dire que la pertinence interne de l'étude n'est pas mise en cause, si la dissymétrie n'est pas attribuée à l'attrition sélective.

Ces limites étant posées, plusieurs recommandations pour des recherches plus approfondies sont proposées. La première est de valider le concept d'agentivité, qui exige d'autres études qui mesurent la perception d'agentivité d'autres produits technologiques, de manière à vérifier si ce concept tient aussi la route dans d'autres domaines. Ces recherches pourraient par exemple étudier la perception de l'agentivité de robots d'assistance aux personnes âgées, de jouets «intelligents» pour les enfants, ou d'ordinateurs embarqués pour véhicules autonomes. La seconde recommandation est de mener l'étude finale avec les mêmes scénarios sur un échantillon représentatif comprenant seulement des civils aux Pays-Bas. Ceci nous permettrait de voir sur quelles valeurs les résultats de l'échantillon militaire et les réponses des civils diffèrent; même si l'on ne peut faire des comparaisons directes en raison du fait que l'échantillon militaire n'est pas représentatif, nous verrions mieux les perceptions des militaires et des civils. Enfin, nous recommandons de mettre en œuvre la «machine morale pour les armes autonomes», décrite dans la section 5, pour donner aux résultats une portée générale, pour lesquels l'étude requière beaucoup plus de personnes interrogées représentant un groupe démographique plus large. Passer à une expérience en ligne à grande échelle, comme la «machine morale», fournirait de grandes quantités de données de groupes démographiques différents qui pourraient être utilisées pour présenter des résultats plus solides et de portée générale, et pour comprendre de manière exhaustive le jugement moral des personnes sur les armes autonomes.

Préface

L'éthique des armes autonomes a commencé à m'intéresser après que j'eus assisté l'Université d'Eté sur l'Intelligence Artificielle Responsable (Al) en Août 2016 organisée par le Dr. Virginia Dignum,[2] sous l'égide de la Conférence Européenne «Intelligence Artificielle» (ECAI). Après mon premier entretien avec Virginia au sujet de mon projet de mémoire, lorsqu'elle m'a demandé si je voulais étudier à l'étranger et en quel endroit, j'ai répondu hardiment «le MIT», mais je n'imaginais pas que cela se concrétiserait. Les relations entretenues par Virginia avec le Dr. Iyad Rahwan, du Groupe de Coopération Evolutive au Laboratoire Media du MIT, ont permis de réaliser ce souhait. Je veux remercier Virginia d'avoir rendu les choses possibles, et de sa supervision tout au cours de ce parcours diplômant. Du fait du décalage horaire, nous avons organisé nos sessions par Skype durant les heures du soir aux Pays-Bas, et je lui adresse tous mes remerciements pour avoir trouvé le temps de revoir mon texte.

Le fait de travailler sur mon mémoire dans le cadre du Groupe de Coopération Evolutive constitua pour moi une énorme opportunité, et le Media Lab constitue un environnement très ouvert, diversifié, sans hiérarchie et très motivant. Au début, il a fallu que je m'habitue à courir dans une «piscine à boules» (ballenbak) dans le couloir, et aux gens qui quittaient leurs bureaux pour jouer au tennis de table en milieu de journée, mais après un mois environ, je me sentais chez moi dans le Lab. J'ai travaillé avec des étudiants très doués

[2] (Note du traducteur) Dr. («Docteur») aux USA fait référence à une personne qui a passé son doctorat quelle que soit sa matière de spécialité, et pas forcément la médecine.

et ambitieux, de deuxième et troisième cycle. J'aimerais en particulier remercier le Dr. Sydney Levine et le Dr. Nick Obradovich de tous leurs commentaires très pertinents; j'ai véritablement apprécié les discussions animées que nous avons menées sur l'organisation de mon travail de recherche. Mais surtout je veux exprimer mes remerciements au Dr. Iyad Rahwan pour m'avoir permis d'être membre de son groupe; j'ai ainsi pu travailler sur le sujet sensible des armes autonomes. Il m'a donné toutes les ressources, le temps et l'assistance dont j'avais besoin pour mener mes recherches, ce dont je suis particulièrement reconnaissante.

Enfin, je voudrais remercier l'Armée de Terre Royale des Pays-Bas de m'avoir permis d'étudier à l'étranger, et de m'avoir soutenue pour mon poste au MIT. Sans son soutien, je n'aurais pas pu vivre et étudier à Boston pendant cinq mois. J'espère que ce mémoire sur l'Ethique des Armes Autonomes fera prendre davantage conscience de ce problème, et conduira à un débat sur le déploiement des armes autonomes et de l'intelligence artificielle dans l'organisation de la Défense.

Ilse Verdiesen

Table

Liste des figures

Liste des tableaux

Introduction

> *Je pense avoir des difficultés à exprimer ce que je ressens à propos des drones. D'un côté, tout instrument de mort me perturbe. D'un autre côté, il est réconfortant de penser que tant de vies humaines ne seront pas mises en danger en les utilisant.*
>
> Pilote interrogé étude 2

Les armes autonomes sont des systèmes d'armes équipés d'intelligence artificielle (IA). L'intelligence artificielle est définie par Neapolitan et Jiang (2012, p. 8) comme étant *«une entité intelligente douée de raisonnement dans un environnement complexe et changeant*», mais cette définition s'applique aussi à l'intelligence naturelle. Russel & Norvig (1995) donnent une liste d'un grand nombre de définitions associant des vues sur des *systèmes qui pensent et agissent comme des humains* et *des systèmes qui pensent et agissent rationnellement*, mais ne donnent pas une définition personnelle claire. Pour le moment, nous adoptons à la définition donnée par Bryson, Kilme et Zürich (2011). Ils avancent qu'une machine (ou un système) fait preuve de comportement intelligent s'il peut choisir un mode d'action fondé sur une observation tirée de son environnement. Dans les publications scientifiques, l'IA est définie comme quelque chose de plus qu'un simple système intelligent. Elle se caractérise par les concepts d'*adaptabilité,* d*'interactivité* et d'*autonomie* (Floridi & Sanders, 2004). Selon Floridi et Sanders, (2004), l'*adaptabilité* signifie que le système peut changer selon son interaction, et peut tirer les leçons de son expérience. Les techniques d'apprentissage des machines en sont un exemple. L'*interactivité* existe lorsque le système et son environnement interagissent, et l'*autonomie* implique que le système peut changer lui-même son état.

Un nombre croissant de chercheurs s'intéresse à la conception responsable de l'IA, qui intègre des valeurs sociales et éthiques, pour prévenir des conséquences sociétales indésirables engendrées par cette technologie. Les principes qui définissent l'IA responsable sont la *capacité à se justifier («accountability»), la responsabilité individuelle («responsibility»)* et *la transparence («transparency»)*. «Accountability» signifie la justification des actions de l'intelligence artificielle, «responsibility» renvoie à la capacité de répondre personnellement de ces actions, et «transparency» est le fait de décrire et de reproduire les décisions que l'IA prend et adapte à son environnement (V. Dignum, 2016).

L'intelligence artificielle n'est pas seulement un scénario futuriste digne des romans de science-fiction dans lesquels des robots humanoïdes, comme Lore le frère de Data dans «Star Trek» ou les Cylons dans «Battlespace Galactica», ont l'ambition de dominer le monde. Un grand nombre d'applications IA sont déjà utilisées aujourd'hui. Les appareils de mesure intelligents, les moteurs de recherche, l'aide à la personne sur les téléphones mobiles, les pilotes automatiques et les automobiles sans chauffeur en sont des illustrations. La présente étude s'intéresse de manière privilégiée à l'application de l'IA dans le domaine militaire, et en particulier au déploiement d'armes autonomes dans un proche avenir. Dans cette introduction, nous exposerons dans un premier temps la problématique de cette étude, puis l'approche de la recherche, la pertinence scientifique et sociétale et l'intégration dans le programme de la filière Architecture de l'Information (Intelligence Artificielle) du Master «Gestion et Stratégie des Technologies des Systèmes» (SEPAM). Pour finir, nous indiquerons brièvement le contenu des chapitres suivants.

1.1 Problématique de recherche

Nous discutons d'abord de l'augmentation du nombre des armes autonomes dans le domaine militaire et mes questions qui s'y rattachent sur l'éthique et l'intelligence artificielle, puis des études sur les drones opérés manuellement dans la section sur l'étude des séquelles. Ensuite, nous définirons notre manque de connaissance, la nature et l'ampleur de ce problème; enfin, les questions majeures et annexes qui se posent à la recherche dans ce domaine.

1.1.1. Analyse de la problématique

Les armes autonomes sont de plus en plus souvent déployées sur les champs de bataille (Roff, 2016). Des études montrent que la Chine possède déjà des véhicules autonomes transportant des robots armés (Lin & Singer, 2014). La Russie prétend avoir à l'étude des chars autonomes (W. Stewart, 2015); les Etats-Unis ont lancé leur premier navire de guerre «sans pilote» en mai 2016 (P. Stewart, 2016), et le fabricant d'armes russe Kalashnikov a récemment révélé qu'il développait un module de combat automatisé qui utilise les réseaux neurologiques (RT, 2017). Les systèmes autonomes peuvent apporter beaucoup de plus-value dans le domaine militaire, par exemple lorsque le pilote automatique d'un F-16 empêche un crash de l'appareil (NOS, 2016), ou par l'utilisation de robots dans les activités de déminage (Carpenter, 2016). Cependant la nature même des armes autonomes peut aussi conduire à des activités incontrôlables et des troubles d'ordre sociétal. On peut en trouver des illustrations dans la campagne «Arrêtons les robots tueurs» suivies par 61 ONG et menée par «Human Rights Watch» (campagne contre les robots tueurs, 2015); également, l'Organisation des Nations Unies exprime ses inquiétudes en affirmant que «*les systèmes d'armes autonomes qui agissent sans contrôle humain significatif devraient être interdits, et la force contrôlée à distance*

ne devrait être employée qu'avec la plus grande prudence» (Assemblée Générale des Nations Unies, 2016). Le déploiement d'armes autonomes sur les champs de bataille sans supervision humaine directe ne constitue pas seulement une révolution dans le domaine militaire selon Kaag et Kaufman (2009), mais peut aussi être considéré comme une révolution morale. Comme le déploiement massif de l'intelligence artificielle sur les champs de bataille semble inévitable (Rosenberg & Markoff, 2016), la recherche sur la responsabilité éthique et morale est impérative.

La prise de décision éthique en matière d'intelligence artificielle et de robots est un nouveau champ d'étude, et plusieurs chercheurs étudient le jugement moral en relation avec ces technologies. Par exemple, Malle (2015) propose un cadre qui associe les domaines (jusqu'à présent) distincts de l'*éthique dans le domaine des robots*, dans lequel les questions éthiques intéressant la conception, le déploiement et le traitement des robots par les humains sont abordées, et la *morale appliquée aux machines*, qui s'intéresse aux questions ayant trait aux capacités morales d'un robot, et à la manière dont elles doivent être mises en œuvre par ordinateur. Cointe, Bonnet et Boissier (2016) proposent un modèle dans lequel un agent peut juger des aspects éthiques de son propre comportement et de celui d'autres agents dans un système multi-agents. Ce modèle décrit un processus de jugement moral qui permet aux agents d'évaluer le comportement d'autres agents. Bonnefon *et al.* (2016) ont étudié les décisions d'ordre éthique qu'un véhicule autonome doit faire, d'ordre utilitaire ou d'autoprotection, lorsqu'il fait face à des piétons sur la route. Dans cette recherche, la «machine morale»[3] au MIT est utilisée pour comprendre en détails la manière dont les personnes voient les scé-

[3] http://moralmachine.mit.edu/

narios mettant en jeu un véhicule autonome, et en quoi leur jugement moral diffère ou non de ceux d'autres personnes.

Dans un domaine ayant rapport avec les armes autonomes, celui des opérations mettant en œuvre des drones pilotés manuellement, les problèmes éthiques ont été étudiés intensivement ces dix dernières années dans des publications de nature philosophique et psychologique. D'un point de vue philosophique, Coeckelbergh (2013) affirme que les opérations mettant en œuvre des drones créent non seulement une distance physique, mais également une distance morale, puisque le visage de l'adversaire devient moins visible, ce qui élimine l'obstacle moral-psychologique et ainsi facilite l'action de tuer. Un autre problème éthique, selon Strawser (2010), est que du fait du changement régulier dans les activités en dehors de la zone de combat, les opérateurs des drones subissent des pressions psychologiques injustes, car ils doivent passer des activités liées au travail aux situations «à la maison» tous les jours. Également, du fait de l'éloignement du champ de bataille, les opérateurs humains peuvent ressentir une dissonance cognitive dans laquelle la guerre est ressentie plus comme un jeu vidéo que comme la réalité. Ces problèmes éthiques sont réfutés par Strawser (2010): il affirme que du fait de l'éloignement les opérateurs humains ont plus le temps d'évaluer une cible, car leur propre sécurité n'est pas menacée, mais il avance également qu'il faudrait procéder à d'autres recherches empiriques pour évaluer les effets psychologiques induits par la conduite d'opérations utilisant des drones à une grande distance du champ de bataille.

Plusieurs études ont été effectuées sur les effets des opérations par drone sur les opérateurs humains dans le domaine de la psychologie. L'une des premières études sur les effets psychologiques de ces opérations a été menée par Thompson *et al.* (2006) qui constatèrent que les opérateurs souffraient

d'une fatigue accrue, d'épuisement émotionnel et de «burnout». Ceci a été confirmé par Chappelle, Goodman, Reardon et Thompson (2014), qui ont fait part du fait que les opérateurs de drones présentent des symptômes de burnout, sont «vidés» émotionnellement, font preuve de beaucoup de cynisme et ont des troubles de type post-traumatique. D'autres facteurs psychologiques mentionnés dans ces études empiriques sont l'ennui et l'inattention du fait de la longue durée des opérations mettant en jeu des drones (Cummings, Mastracchio, Thornburg & Mkrtchyan, 2013). Ces études montrent que le fait d'effectuer des opérations utilisant des drones peut avoir de graves effets psychologiques sur les humains qui actionnent ces drones.

1.1.2 Vide dans les connaissances

Dans le débat sur les armes autonomes des vues et des opinions bien tranchées sont exprimées. La Campagne pour «Arrêter les Robots Tueurs» (2017) déclare par exemple sur son site internet que *«Permettre que des décisions sur la vie ou la mort soient prises par des machines transgresse une limite morale fondamentale. Des robots autonomes sont dépourvus de jugement humain et de la capacité à comprendre un contexte».* Nous avons trouvé peu d'études qui vont dans le même sens ou qui donnent un aperçu de la manière dont les Armes Autonomes sont perçues par le grand public et par les militaires. L'étude «Open Robots Ethics initiative» a effectué un sondage d'opinion en 2015, et publié un rapport. Mais les résultats ne furent pas publiés dans une revue spécialisée, et le sondage n'était pas assez représentatif pour que l'on puisse en tirer des conclusions significatives. Comme on l'a signalé dans le paragraphe précédent, plusieurs spécialistes comme Malle (2015), Cointe *et al.* (2016) et Bonnefon *et al.* (2016) étudient le processus de prise de décision d'ordre éthique dans le domaine de l'intelligence artificielle et des ro-

bots. Les problèmes éthiques sont également étudiés dans le domaine voisin des opérations mettant en œuvre des drones (Coeckelbergh, 2013; Strawser, 2010), et les études empiriques ont montré que les opérateurs souffraient de graves effets psychologiques (Chappelle *et al.* 2013; Thompson *et al.* 2006), mais ce type de recherche n'a pas encore été étendu au déploiement d'armes autonomes. Nous n'avons pas trouvé non plus de publications ni d'études empiriques sur les valeurs morales mises en jeu par les armes autonomes, ni sur ce que certaines personnes considèrent comme «la limite morale fondamentale». Par conséquent, le trou dans les connaissances est double, en ce sens que sont absentes les connaissances précises sur (1) la manière dont les armes autonomes sont perçues par les militaires et le grand public, et (2) les valeurs morales que les militaires et le grand public considèrent importantes lors du déploiement, dans un futur proche, des armes autonomes.

1.1.3 Enoncé de la problématique

La première partie de ce «trou» dans les connaissances peut être comblée en étudiant la perception des armes autonomes, en utilisant la théorie de l'agentivité énoncée dans les domaines de la psychologie cognitive, de l'intelligence artificielle et de la philosophie morale. L'«agentivité» est la capacité de planifier et d'agir (Waytz, Gray, Epley & Wegner, 2010), et est étudiée relativement à des questions touchant ou non l'humain. On attribue de l'agentivité à des objets, comme les ordinateurs (Nass, Moon, Fogg, Reeves & Dryer, 1995) et les robots (H.M. Gray, Gray, & Wegner, 2007); on peut donc penser que la technologie IA sera également perçue comme possédant les caractéristiques de l'agentivité. L'étude des perceptions qu'ont les militaires et le grand public vis-à-vis des armes autonomes permettrait de distinguer des différences et des ressemblances pos-

sibles entre les points de vue de ces groupes, qui pourraient être utilisées dans le débat actuel sur cette technologie.

La deuxième partie peut être comblée en étudiant les théories des valeurs connues, pour déterminer quelles sont les valeurs que l'on considère comme importantes concernant le déploiement des armes autonomes. Les théories sur les valeurs les plus reconnues sont celles de Schwartz (1994), Friedman & Kahn Jr. (2003) et Beauchamp & Walters (1999), mais on ne voit pas comment on peut les appliquer aux armes autonomes. Le fait de dégager les valeurs les plus pertinentes dans le contexte du déploiement des armes autonomes, et de les comparer avec celles qui sont mises en jeu dans la technologie actuelle des drones actionnés par l'homme, conduira à percevoir ce qui est sous-jacent dans le débat sur les armes autonomes, et permettra une meilleure compréhension des vues exprimées.

Ceci nous amène à l'énoncé de notre problématique:

> *Dans le débat sur les armes autonomes, des vues et des opinions bien tranchées sont exprimées, mais la recherche empirique qui sous-tend ces opinions est absente. Nous ne savons pas comment les armes autonomes sont perçues, ni quelles sont les valeurs morales qui sont mises en jeu dans le déploiement des armes autonomes prévu dans un avenir proche. On ne voit pas non plus clairement si la perception et les valeurs morales relatives aux armes autonomes qu'ont les personnels militaires est différente de celle du grand public.*

1.1.4. Champ de recherche

Beaucoup de publications sur les armes autonomes sont écrites et discutées par des experts juridiques et des philosophes dans le contexte du droit humanitaire international et les Conventions de Genève, dont le but est de limiter les effets des conflits armés (CICR). Comme nous ne sommes ni des experts juridiques ni des philosophes, la présente étude restera à

l'intérieur des limites et des lois des conflits armés, et nous ne les remettrons pas en question. Du fait des contraintes imposées par les délais, la présente étude portera uniquement sur les personnels du Ministère de la Défense des Pays-Bas, et ne sera pas élargie à un échantillon représentatif de la population civile néerlandaise. Dans cette étude, nous nous intéresserons au déploiement des armes autonomes dans un avenir proche, que nous définissons comme *les cinq prochaines années*. Ceci a pour conséquence que nous n'étudierons pas les armes équipées d'une «intelligence artificielle générale», ni la technologie futuriste qu'il n'est pas encore possible de décrire; par contre, nous nous intéresserons à la technologie qui est développée actuellement, spécifiquement les drones dotés d'une capacité de ciblage autonome. Le processus de ciblage s'articule en six étapes: (1) trouver, (2) fixer l'objectif, (3) poursuivre l'objectif (4) cibler, (5) prendre à partie, (6) évaluer. Nous traiterons de la prise de décision aux étapes 4 et 5 du processus de ciblage, du fait que de nombreuses décisions d'ordre éthique doivent être prises à ce moment du processus (Asaro, 2016).

1.1.5. Question posée par cette étude, et questions secondaires

En se basant sur le manque de connaissances qui a été identifié, sur l'énoncé de la problématique et sur le champ de recherche, nous proposons d'énoncer la question à laquelle tente de répondre la présente étude:

> *Comment les armes autonomes sont-elles perçues par le grand public et par les personnels militaires du Ministère de la Défense néerlandais, et quelles valeurs morales mises en jeu par le déploiement d'armes autonomes considèrent-ils comme importantes?*

Pour répondre à cette question fondamentale, on examinera les questions secondaires suivantes:

1. *Comment les armes autonomes sont-elles définies dans les publications existantes?*
2. *Quelles théories des valeurs sont-elles exposées dans les publications existantes?*
3. *Quelles sont les valeurs exposées dans les théories des valeurs qui sont mises en jeu par les armes autonomes?*
4. *De quelle manière l'agentivité est-elle définie dans les publications existantes?*

Pour répondre à ces quatre questions secondaires, on se servira des publications existantes examinées dans cette étude.

5. *Quelles valeurs morales sont-elles mises en relation par le grand public et les personnels militaires du Ministère de la Défense néerlandais avec le déploiement d'armes autonomes?*

Pour répondre à cette question secondaire, on procédera à une enquête sur les valeurs consistant en un questionnaire mis en ligne et en des entretiens avec des spécialistes.

6. *Comment l'agentivité des armes autonomes est-elle perçue par le grand public et par les personnels militaires du Ministère de la Défense néerlandais?*
7. *De quelle manière les valeurs mises en jeu par les armes autonomes sont-elles perçues par le grand public et par les personnels militaires du Ministère de la Défense néerlandais?*

Pour répondre à ces deux questions secondaires, on réfléchira sur une hypothèse et une analyse descriptive basées sur les résultats d'une expérience contrôlée randomisée.

8. *Comment les valeurs mises en jeu par les armes autonomes peuvent-elles être intégrées dans la conception d'une «machine morale» pour les armes autonomes?*

On peut répondre à cette question secondaire en créant un plan de conception et de mise en œuvre d'une «machine morale» pour les armes autonomes.

Dans les paragraphes suivants, l'approche de recherche pour répondre à ces sous-questions est décrite plus en détails.

1.2. Approche de recherche

Dans la présente étude, nous employons la méthode de la conception éthique (conception intégrant les valeurs / «value-sensitive design» ou VSD) comme approche. La VSD est une approche à trois volets qui permet de tenir compte des valeurs humaines tout au long de la conception technologique. C'est un processus itératif intéressant l'investigation des valeurs humaines dans les champs *empirique, conceptuel et technologique* mises en jeu dans la conception (Davis & Nathan, 2015; Friedman & Kahn Jr, 2003). L'investigation conceptuelle consiste en deux volets: (1) identification des acteurs directs, ceux qui utiliseront cette technologie, et les acteurs indirects, ceux dont la vie est affectée par cette technologie, et (2) identification et détermination des valeurs que l'emploi de cette technologie met en jeu. L'investigation empirique s'intéresse à la compréhension et à l'expérience des acteurs dans un contexte associé à cette technologie, et les valeurs mises en jeu seront examinées. Dans l'investigation technique, les caractéristiques spécifiques de cette technologie sont analysés (Davis & Nathan, 2015). La conception éthique peut être utilisée comme «feuille de route» par les ingénieurs et les étudiants pour intégrer des considérations éthiques dans la conception (Cummings, 2006).

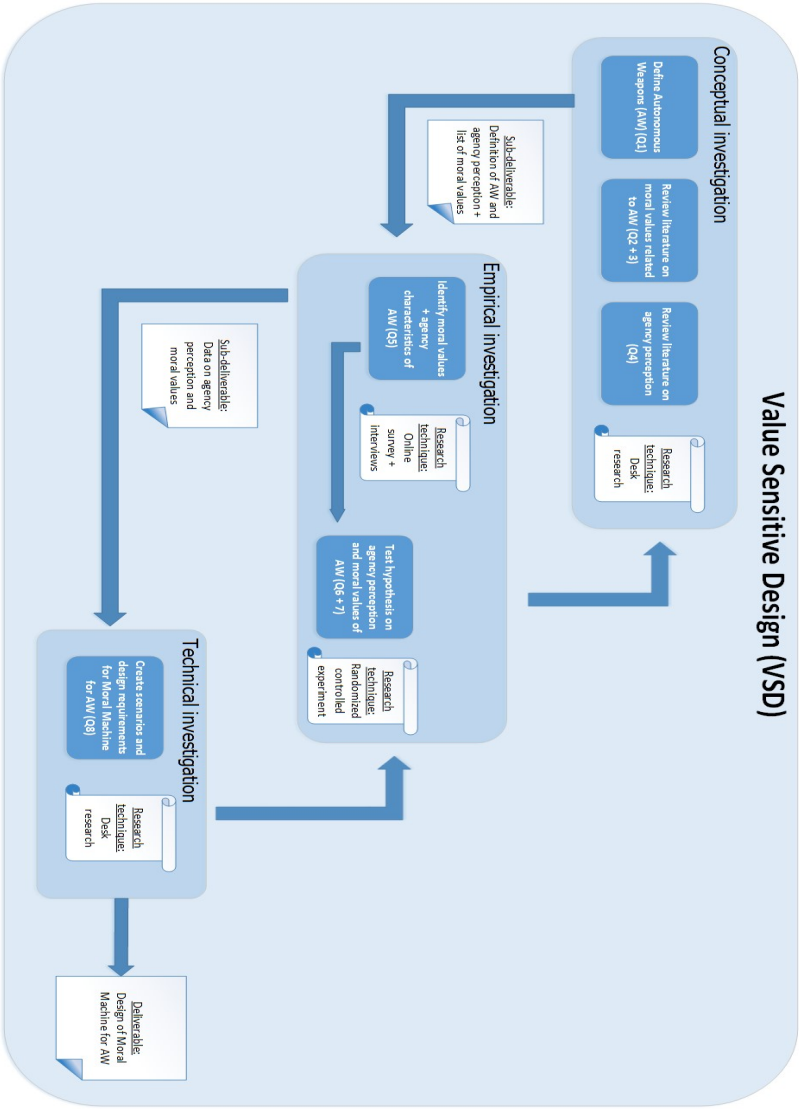
Quelques critiques ont également été exprimées concernant l'approche conception éthique. L'un des problèmes mentionnés par Davis et Nathan (2015) est que la conception éthique tient pour acquis que certaines valeurs sont universelles, mais qu'elles peuvent varier selon les cultures et le contexte. Pour réfuter cet argument, on pourrait prendre une base empirique comme point de vue de départ plutôt qu'une base philosophique, ou reconnaître que la position des chercheurs n'est pas forcément la seule valide (Borning & Muller, 2012). Borning et Muller (2012) avancent une proposition pluraliste selon laquelle la conception éthique ne doit pas recommander

d'opinions sur les valeurs, universelles ou relatives, mais doit laisser aux ingénieurs la liberté de décider quelle opinion est la plus appropriée dans le contexte de la conception.

En accord avec Borning et Muller (2012) nous avons suivi l'approche basée sur la conception éthique (VSD) comme guide de recherche et non comme fin en soi. Pour le volet conceptuel, nous dévions légèrement de la méthode VSD originale, car nous ne procédons pas à une analyse complète pour identifier les acteurs dans la phase 1, mais nous nous concentrons sur deux groupes d'acteurs évident: *le grand public* et *les personnels militaires*; en effet, ces deux groupes étaient disponibles pour l'enquête. Dans la deuxième étape de la première phase, nous exposons une revue des publications existantes dans le but d'identifier les valeurs mises en jeu par les armes autonomes. Le principal centre d'intérêt porte sur la phase d'investigation empirique, car nous avons constaté que l'aspect empirique est négligé dans le débat éthique sur les armes autonomes. Nous nous sommes tournés vers les critiques de Davis et Nathan (2015) en consacrant la plus grande partie de notre temps à cette phase de la VSD, dans laquelle nous étudions les cas où les armes autonomes utilisent diverses techniques de recherche. Dans la phase d'investigation technique, nous ne concevons pas une arme autonome de la manière dont on pourrait s'y attendre de façon intuitive, car cela constituerait un projet immense dépassant de beaucoup le propos de la présente étude. Néanmoins, nous avons décidé d'exploiter le travail du groupe de recherche «Coopération évolutive» (Scalable Cooperation research group), et nous proposons une conception pour une «machine morale» destinée aux armes autonomes qui pourrait être utilisée dans une prochaine phase de la recherche portant sur l'éthique des armes autonomes. Nous appliquons les phases de la VSD à notre travail de recherche de la façon suivante (Figure 1):

- **Investigation conceptuelle:** La pénurie de travaux de recherche sur les armes autonomes implique que la première phase de notre étude est par nature exploratoire. Les activités de recherche concernant l'investigation conceptuelle seront effectuées qualitativement sur le mode de la recherche documentaire, pour produire une revue des publications existantes sur les armes autonomes, les valeurs morales et la perception de l'agentivité.
- **Investigation empirique**: Cette phase s'articule en deux parties. La première phase est exploratoire, et identifie les valeurs dans le contexte des armes autonomes au moyen d'une enquête en ligne qui constitue une technique de recherche quantitative, complétée par une technique de recherche qualitative: des entretiens avec des experts. Dans la seconde partie de l'investigation empirique, nous utilisons à la fois une méthode de recherche exploratoire et confirmative pour étudier la perception de l'agentivité. Nous opérationnalisons le concept d'agentivité pour confirmer notre hypothèse sur la perception de l'agentivité en menant des expériences contrôlées randomisées, et utilisons les résultats de l'expérience pour explorer les valeurs mises en jeu par les armes autonomes, et ce de façon descriptive.
- **Investigation technique**: Dans cette phase, la conception et les caractéristiques de la «machine morale» destinée aux armes autonomes sont produites à partir de scénarios qui sont utilisés dans l'expérience contrôlée randomisée de la phase empirique. La phase d'investigation technique est exploratoire, et le concept est créé en utilisant la technique de recherche documentaire.

Figure 1 – L'approche de recherche sur la conception éthique



1.3. Pertinence / applicabilité

Dans les paragraphes qui suivent, nous définirons la pertinence scientifique et sociétale de cette étude, et montrerons comment celle-ci s'intègre dans la démarche «intelligence artificielle» du master Politique et Gestion de l'Ingénierie des Systèmes (System Engineering Policy and Management, SEPAM)

1.3.1. Pertinence scientifique

Cette étude a une pertinence scientifique dans la mesure où elle contribue à la «littérature» scientifique en donnant un aperçu de la perception qu'ont le grand public et les militaires des armes autonomes, et en identifiant les valeurs morales que le grand public et les militaires associent aux armes autonomes. Actuellement nous manquons de connaissances en la matière, et aucune donnée sur la perception et les valeurs associées aux armes autonomes n'a été trouvée. En utilisant la conception éthique (VSD) comme approche de recherche, nous montrerons qu'il est possible de l'utiliser en tant que structure de recherche académique, alors qu'elle pourrait être considérée par certains comme méthode d'approche de type VSD pour des études de cas uniquement. Également, nous élargissons les travaux de recherche sur la prise de décision éthique concernant les véhicules autonomes (qui ont été menés par Bonnefon *et al.* en 2016) au domaine des armes autonomes.

1.3.2. Pertinence d'ordre sociétal

La pertinence sociétale de cette étude est que la compréhension de la perception qu'ont le grand public et les personnels militaires du Ministère de la Défense néerlandais, et que l'identification des valeurs morales qu'ils associent aux armes autonomes, peuvent être utiles pour définir les point d'accord et de désaccord dans le débat ayant trait à cette technologie et lancé par la Campagne pour Arrêter les Robot Tueurs (2015,

Campaign to Stop Killer Robots) et le Comité International pour le Contrôle des Armes Robotisées (International Committee for Robot Arms Control (ICRAC). En second lieu, les résultats de cette étude démontrent que la méthode de conception éthique peut être appliquée aux armes autonomes pour identifier les valeurs que les militaires et le grand public associent au déploiement de ce type d'armes. Enfin, en identifiant les valeurs qu'il est important d'intégrer dans la conception des armes autonomes, la présente étude peut contribuer à promouvoir une conception et un déploiement responsables des armes autonomes dans le futur.

1.3.3. Intégration dans le programme SEPAM et dans la démarche Intelligence Artificielle

Les critères généraux dans le cadre d'un mémoire de master retenus à la faculté de technologie, stratégie et gestion des systèmes sont les suivants: l'étude doit avoir une composante analytique, se concentrer sur un domaine technique et être de nature multidisciplinaire. Pour obtenir son diplôme de master SEPAM, il faut que le travail de recherche échafaude une solution à un problème complexe, important, contemporain, de nature sociotechnique, ce qui signifie que le mémoire doit se focaliser clairement sur un domaine technique impliquant différents acteurs, prendre en compte les valeurs privées et publiques, ne traite pas de problèmes uniquement techniques mais aussi de choix en matière de gestion et d'éthique (Portail diplômes, 2017). La filière Architecture de l'Information intègre les volets gestion et informatique, et se concentre sur la compatibilité des besoins organisationnels et des opportunités en matière d'ingénierie pour trouver des solutions TOC (technologies de l'information et de la communication) à la pointe du progrès (Programme Architecture de l'Information, 2017).

La présente étude respecte les critères SEPAM en ce qu'elle se concentre spécifiquement sur la technologie IA des armes autonomes dans le domaine militaire. Elle a un caractère multidisciplinaire puisqu'elle associe les publications sur la perception de l'agentivité dans les domaines de la psychologie cognitive, de l'intelligence artificielle et de la philosophie morale, de la théorie des valeurs définie dans les textes de nature philosophique et psychologique, et le déploiement d'armes dans le domaine militaire. Le problème fondamental de nature sociotechnique dans la présente étude est le manque actuel de connaissances pour ce qui est de la perception qu'ont le grand public et les militaires de la technologie des armes autonomes, et des valeurs qu'ils jugent importantes dans ce domaine. Cette étude respecte l'optique Intelligence Artificielle, car elle identifie les points d'accord et de désaccord dans les priorités d'ordre éthique des militaires et du grand public. Elle propose une solution d'ingénierie pour la conception d'une «machine morale» destinée aux armes autonomes, de manière à explorer plus précisément ces priorités au moyen d'une enquête en ligne de grande ampleur, utilisant le concept de «machine morale» pour les véhicules autonomes.

1.4. Structure

Le reste de cette étude est organisé comme suit: dans la section 2, les publications sur les armes autonomes, les théories des valeurs et la perception de l'agentivité sont présentées. La section 3 décrit la méthodologie, les hypothèses, la conception de recherche, l'opérationnalisation des scénarios et des concepts, l'approche analytique, le pré-enregistrement, l'échantillon, et conclut avec les questions méthodologiques. Dans la section 4, tout d'abord les résultats de l'enquête sur les valeurs sous forme de questionnaire en ligne et d'entretiens de recherche sont donnés, suivis des résultats des trois expériences contrô-

lées randomisées. Le concept de la «machine morale» pour les armes autonomes est présenté dans la section 5, tout d'abord la «machine morale» originale pour les véhicules autonomes, puis les caractéristiques et les scénarios pour intégrer une «machine morale» dans des armes autonomes, et un plan de mise en œuvre succinct. La section 6 conclut sur les résultats, discute des implications d'ordre scientifique et sociétal, et identifie les limites et les recommandations en vue d'études ultérieures. Pour finir, dans la section 7, nous faisons part de réflexions sur les choix auxquels nous avons procédé dans notre recherche, la construction de ce projet, et le lien entre la recherche, l'intelligence artificielle et le programme SEPAM.

2. Revue des publications

> *C'est simplement que je ne me sens pas à l'aise devant le fait de placer des armes sur une machine qui n'a pas d'émotions et qui n'est pas contrôlée par l'homme.*
>
> Pilote interrogé, étude 1

Cette section donne une présentation des publications existantes ayant trait aux armes autonomes, suivie d'un résumé des théories des valeurs, et enfin nous définissons le concept d'agentivité que l'on utilise pour étudier la perception des armes autonomes. Chaque sous-section se termine sur une discussion des théories que l'on applique dans cette étude, ainsi que des raisons de leur choix.

2.1. Les armes autonomes

Bien que le débat sur les armes autonomes ait attiré beaucoup l'attention ces dernières années, nous avons trouvé que le sujet n'était pas bien cerné dans les publications académiques. Dans cette sous-section, nous commencerons avec une revue des nombreuses et diverses définitions, et présenterons deux classifications des armes autonomes pour conclure cette section.

2.1.1. Définition

Les armes autonomes sont une technologie «émergente», et il n'en existe pas encore de définition reconnue sur le plan international (AIV & CAVV, 2016). Il n'y a même pas consensus sur le fait de savoir s'il y a lieu d'en donner une définition. Bien que quelques auteurs donnent des définitions dans leurs écrits (tableau 1), d'autres mettent en garde contre de telles précisions. *Pour l'OTAN, «on doit éviter d'essayer de créer des définitions pour les «systèmes autonomes» car, par définition, les machines ne peuvent pas être*

autonomes au sens littéral du terme». (Kuptel & Williams, p. 10). L'Institut de la Recherche pour le Désarmement des Nations Unies émet également des doutes sur le fait de donner une définition des armes autonomes; ses membres avancent que le degré d'autonomie dépend des «fonctions critiques considérées et des interactions de différentes variables» (UNIDIR, 2014, p. 5). Ils affirment que l'une des raisons de la différence entre les termes concernant les armes autonomes est que parfois les choses (drones ou robots) sont définies, mais parfois aussi une caractéristique (l'autonomie), des variables dans les problèmes (létalité ou degré de contrôle humain), ou la pratique (ciblage ou mesures défensives) apparaissent dans la discussion et deviennent une partie de la définition.

Tableau 1 Aperçu des différentes définitions des armes autonomes

Auteur	Définition
AIV & CAVV (2016, p.11)	*«Arme qui, sans intervention humaine, choisit en prend à partie des cibles qui correspondent à certains critères prédéfinis, à la suite d'une décision d'origine humaine de déployer l'arme en sachant qu'une attaque, une fois lancée, ne peut être arrêtée par une intervention humaine».*
Altmann, Asaro, Sharkey & Sparrow (2013, p. 73)	Les armes autonomes sont: *"des armes robotisées qui, une fois actionnées, choisissent et vont frapper des cibles sans autre intervention humaine"*
Galliott (2015, p. 5)	Les robots militaires sont: *«un groupe de systèmes électromécaniques propulsés, qui ont tous les points communs suivants:* *1. Ils n'ont pas de personnels embarqués;* *2. Ils sont conçus pour être récupérables (même s'ils peuvent ne pas être utilisés dans ce but); et*

	3. Dans un contexte militaire, ils peuvent utiliser leurs capacités pour délivrer une charge létale ou non létale, ou avoir une fonction d'appui pour des objectifs d'une force militaire».
Horowitz (2016, p. 27)	*«un système d'arme qui, une fois actionné, est destiné à frapper seulement des cibles individuelles ou des groupes de cibles spécifiques qui ont été choisis par un opérateur humain».*
Royakkers & Orbons (2015, p. 625)	Les robots militaires sont: *«… des systèmes sans personnels embarqués réutilisables destinés à usage militaire, doté d'un degré d'autonomie variable».*
Kuptel & Williams (2014, p. 10)	*«Les machines ne sont «autonomes» que dans certaines fonctions, comme la navigation, l'optimisation des capteurs ou la gestion de la consommation de carburant».*
UNDIR (2014, p. 5)	Le degré d'autonomie dépend des *«fonctions essentielles et de l'interaction de différentes variables».*

Les différentes définitions des armes autonomes sont données dans le Tableau 1. Certains auteurs utilisent le terme «robots militaires» lorsqu'il existe un certain degré d'autonomie. Comme les robots militaires peuvent être considérés comme une sous-catégorie des armes autonomes selon la classification de Royakkers & Orbons (2015) (Figure 2), nous les avons intégrés dans la liste des définitions. Selon nous, la définition donnée dans le rapport du Conseil Consultatif des Affaires Internationales (AIV & CAVV) saisit au mieux la description des armes autonomes d'un point de vue militaire et d'ingénierie, car elle tient compte de critères prédéfinis, et aussi du processus militaire de ciblage, puisque l'arme ne sera déployée qu'après prise de décision humaine. Par conséquent, nous adopterons cette définition, et définirons une arme autonome comme:

> *«Une arme qui, sans intervention humaine, choisit et prend à partie des cibles qui correspondent à certains critères prédéfinis, à la suite d'une décision d'origine humaine de déployer l'arme en sachant qu'une attaque, une fois lancée, ne peut être arrêtée par une intervention humaine».* (AIV & CAV, 2016, p. 11).

2.1.2. Classification des armes autonomes

Les armes autonomes sont non seulement définies de manière ambigüe, mais elles n'ont pas été classifiée de manière uniforme. Nous présentons deux de ces classifications dans cette sous-section. Royakkers & Orbons (2015) définissent plusieurs types d'armes autonomes (figure 2), faisant la distinction entre (1) *les armes non létales*, qui sont des armes «*ne faisant pas de victimes (innocentes) ou ne nuisant pas de manière grave et permanente aux personnes».* (Royakkers &Orbons, 2015, p. 617), comme le système d'interdiction active (Active Denial System) qui met en œuvre un rayon d'énergie électromagnétique pour maintenir des personnes à une certaine distance d'un objet ou de troupes, et (2) les *Robots Militaires*, qu'ils définissent comme étant *«des systèmes sans personnels embarqués réutilisables destinés à un usage militaire, dotés d'un degré d'autonomie variable*» (Royakkers & Orbons, 2015, p. 625). Les robots militaires peuvent être classés en trois catégories: les véhicules basés à terre, par exemple destinés à des missions de reconnaissance sans personnels embarqués et à neutraliser les engins explosifs improvisés (IED), les véhicule pouvant circuler sans personnels embarqués sous l'eau ou à la surface de l'eau, comme une pièce d'artillerie placée sur un navire ou un sous-marin autonome, ou des aérodynes de combat télépilotés (unmanned combat aerial vehicles, UCAV). Ces aérodynes de combat sont classifiés par Royakkers et Orbons (2015) comme étant télé-actionnés; les «drones» en sont l'exemple le plus connu, ainsi que les UCAV autonomes, qui sont peu à peu

développés par le Département de la Défense des Etats-Unis (Rosenberg & Markoff, 2016).

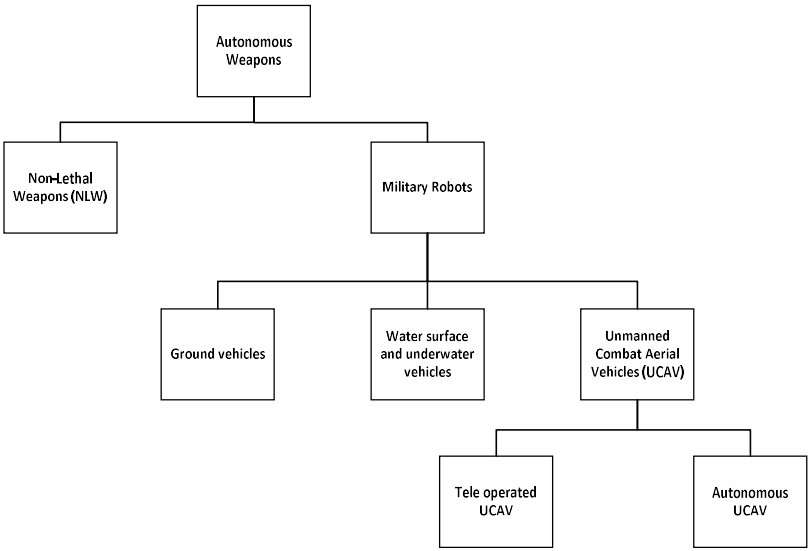


Figure 2 Classification des armes autonomes selon Royakkkers & Orbons (2015)

Galliot (2015) propose un autre type de classification des armes autonomes, basé sur quatre niveaux d'autonomie pour les systèmes sans équipage (sans personnels embarqués):

1. NIVEAU D'AUTONOMIE 1 – NON-AUTONOME / OPERE A DISTANCE:

«*Un opérateur humain contrôle chacun des mouvements mécaniques sans exception. Sans opérateur, les systèmes actionnés à distance sont incapables d'effectuer une quelconque opération effective*».

2. NIVEAU D'AUTONOMIE 2 – AUTONOMIE SUPERVISEE:
«Un opérateur humain précise les mouvements, les positions et les actions simples, et le système effectue ces tâches. L'opérateur doit fréquemment fournir des instructions au système et effectuer une supervision diligente pour s'assurer que les actions sont menées correctement».

3. NIVEAU D'AUTONOMIE 3 – AUTONOMIE DANS L'ACTION:
«*Un opérateur humain spécifie une tâche générale, et le système élabore un mode d'action et le mène à bien sous son propre contrôle. Typiquement, l'opérateur a les moyens de surveiller le système, mais cela n'est pas nécessaire pour l'action à mener».*

4. NIVEAU D'AUTONOMIE 4 – AUTONOMIE TOTALE:
«Un système doté d'une autonomie totale crée et mène à bonne fin ses propres tâches sans avoir besoin d'instructions d'origine humaine, à l'exception de la décision d'élaborer ce système. L'agent humain est tellement éloigné du processus que le niveau d'influence directe est négligeable. Ces systèmes pourraient faire preuve de capacités qui imitent ou dupliquent les capacités morales des êtres humains dotés de sentiments (bien que nous ne prendrons pas position ici sur cette question)» (Galliott, 2015, p. 7).

Cette classification est à notre avis une bonne démarche pour catégoriser le niveau d'autonomie des armes autonomes, mais nous émettons quelques réserves d'un point de vue de l'ingénierie. Galliott (2015) lui-même déclare qu'il serait possible de fusionner les deuxième et troisième niveaux d'autonomie, car ils sont tous deux des niveaux opérationnels semi-autonomes. Nous sommes d'accord avec cet argument, mais ceci n'est pas le problème principal que nous rencontrons dans ces définitions. Nous pensons qu'il est curieux de commencer une liste de niveaux d'autonomie par la catégorie des systèmes non autonomes. Plus important: dans le quatrième

degré d'autonomie, l'auteur déclare que «*ces systèmes pourraient faire preuve de capacités qui imitent ou dupliquent les capacités morales des êtres humains dotés de sentiments*». Il semble se référer à la définition de l'intelligence artificielle forte, pour laquelle un ordinateur a des états et des programmes cognitifs pouvant imiter la connaissance humaine (Searle, 1980). Affirmer qu'un système autonome est doté de capacités morales montre à notre avis un manque de connaissances sur les systèmes d'intelligence artificielle actuels, ceux-ci n'étant pas autre chose que des ordinateurs ayant des capacités d'interactivité, d'autonomie et d'adaptabilité (Floridi & Sanders, 2004).

Comme il n'est pas sûr qu'une IA forte et capable de «capacités morales propres aux humains dotés de sentiments» soit un jour développée, nous pensons que la classification de Galliott (2015) n'est pas réaliste au vu de l'état de la technologie actuelle; par conséquent elle ne sera pas utilisée dans la présente étude. Pour cette dernière, nous adoptons la classification de Royakkers et Orbons (2015) qui donne un bon aperçu de la technologie militaire d'aujourd'hui et de demain. En particulier, leur distinction entre les aérodynes de combat sans équipage télépilotés d'une part et autonomes d'autre part est réaliste, et convient à notre travail de recherche.

2.2. Valeurs

Contrairement au problème des armes autonomes, le concept des valeurs a été étudié en détail dans les domaines de la philosophie et de la psychologie. Cette section présente une définition des valeurs, suivie d'une vue d'ensemble des théories qui définissent les valeurs universelles, une vue d'ensembles des valeurs mises en jeu par les armes autonomes, et conclut sur une hiérarchie des valeurs pour donner un exemple afin de relier les phases d'investigation conceptuelle et empirique de notre recherche.

2.2.1. Définition

Les théories des valeurs ont été bien étudiées dans les domaines de la philosophie morale et de la psychologie. La philosophie morale a une longue et riche histoire pour ce qui est de l'examen des valeurs, et dans ce domaine des questions d'ordre théoriques sont posées pour étudier la nature de la valeur et du bien (Schroeder, 2016). Une distinction est souvent faite entre les valeurs *instrumentales*, qu'il est raisonnable de privilégier pour leurs effets sur ce qui est bien (Ronnow – Rasmussen, 2002), et les valeurs *intrinsèques*, qui *«sont une catégorie de valeurs qui, une fois acquises, sont acquises uniquement en vertu de leurs propriétés intrinsèques»* (Bradley, 2006, p. 112). Bien que la philosophie morale s'intéresse surtout aux théories de ce qui *«devrait être»*, et qu'elle ne soit pas au sens strict influencée par des résultats empiriques (Alfano & Loeb, 2014), elle comporte une branche qui intéresse l'éthique appliquée, qui entre dans le champ de notre étude, car l'éthique appliquée établit le lien entre les théories éthiques abstraites et la pratique morale. Comme il a été dit dans la section 1.2., le propos de la présente étude est la phase empirique de la conception éthique pour rechercher comment les valeurs morales du grand public et des militaires coïncident ou non avec les armes autonomes. Par conséquent, nous avons décidé de ne pas utiliser les théories des valeurs de la philosophie morale dans notre étude; nous nous sommes tournés vers le domaine de la psychologie et de l'éthique appliquée pour obtenir une vue empirique des valeurs personnelles des deux groupes de parties prenantes sélectionnés, de manière à obtenir un aperçu de la situation telle qu'elle est et non telle qu'elle «devrait» être.

Le domaine de la psychologie distingue les valeurs des attitudes, besoins, normes et comportement en ce qu'elles sont une croyance (ou conviction), mènent au comportement qui guide les personnes, et sont ordonnées à l'intérieur d'une hié-

rarchie qui montre l'importance d'une valeur par rapport à d'autres valeurs (Schwartz, 1994). Les valeurs sont utilisées par les personnes pour justifier leur conduite, et définissent quels types de comportement sont socialement acceptables (Schwartz, 2012). Elles sont distinctes des faits en ce que les valeurs ne décrivent pas seulement un bilan empirique du monde extérieur, mais aussi adhèrent aux intérêts des humains dans un contexte culturel (Friedman *et al*, 2013). Les valeurs peuvent être invoquées pour expliquer la prise de décision individuelle, et pour étudier la dynamique humaine et sociale (Cheng & Fleischmann, 2010).

De nombreuses définitions des valeurs existent. Par exemple, Schwartz (1994, p. 21) définit les valeurs comme *des buts transituationnels souhaitables, d'importance variable, servant de principes directeurs dans la vie d'une personne ou d'une autre entité sociale».* Ceci constitue une définition précise, en comparaison de Friedman, Kahn JR, Borning et Huldtgren (2013, p. 57) qui définissent les valeurs comme étant *«ce qu'une personne ou un groupe de personnes considèrent comme important dans la vie».* Les définitions existantes ont été résumées par Cheng et Fleischmann (2010, p. 2) dans leur méta-inventaire des valeurs; ils avancent que *«les valeurs servent de principes directeurs à ce que les gens considèrent important dans la vie».* Bien que ce soit une définition tout à fait simple, nous pensons qu'elle reflète au mieux la description d'une valeur, car elle combine plusieurs définitions en une seule, en utilisant les principales caractéristiques des valeurs. Par conséquent, nous adopterons la définition de Cheng et Fleischmann (2010) pour la présente étude.

2.2.2. Les valeurs universelles

Les travaux de recherche montrent que les personnes, dans toutes les cultures, s'identifient à des valeurs fondamentales qui peuvent être considérées comme des valeurs humaines univer-

selles (Friedman *et al*, 2012; Schwartz, 2012). Bien qu'il y ait des différences selon les individus dans l'importance que l'on attribue aux valeurs, il semble qu'il existe un degré de consensus étonnamment élevé à travers les cultures sur l'ordre hiérarchique des valeurs (Schwartz, 2012). Quelques auteurs consacrent une partie de leur recherche à faire ce que l'on pourrait appeler un inventaire des valeurs, qui sont des listes qui peuvent être utilisées pour catégoriser l'analyse des valeurs humaines, et qui sont souvent accompagnées d'un outil descriptif permettant la discussion sur ces valeurs (Cheng & Fleischmann, 2010). Les inventaires de valeurs les plus courants et les plus documentés sont ceux de Schwartz (1994), Friedman *et al* (2013), Beauchamp & Walters (1999), et Graham *et al* (2012). Le nombre de valeurs universelles trouvées par les chercheurs varie considérablement. Un aperçu de ces inventaires de valeurs est donné dans le Tableau 2, et les théories seront brièvement exposées dans le paragraphe qui suit.

Après avoir procédé à des recherches poussées, Schwartz (1994) mentionne 10 types de valeurs incitatives, divisés en une liste plus fine de 56 valeurs qu'il utilise pour examiner les 10 valeurs universelles de portée générale. Dans leur description de leur approche de la conception éthique, Friedman *et al* (2013) mentionnent 12 valeurs; les 9 premières sont basées sur des orientations morales conséquentialistes et déontologiques, et les 3 dernières sont choisies à partir du champ de l'Interaction Humaine Informatisée (HCI). Graham *et al* (2012) utilisent le terme «fondation» pour définir les 5 valeurs distinctes spécifiques à l'universalité de la nature morale humaine, que Haidt et Joseph (2004) utilisent comme base de la Théorie de la Fondation Morale. Gouveia, Milfont et Guerra (2014) ont élaboré un cadre basé sur de nombreuses théories des valeurs, comme la hiérarchie des besoins de Schwartz (1994) et Maslow (1943). Dans ce cadre, les auteurs situent la valeur sur deux di-

mensions: (1) avec les actions qui motivent le comportement humain qui peuvent être des buts personnels, centraux ou sociaux; et (2) des incitateurs qui représentent les besoins humains, qui peuvent être divisés en besoins de prospérité (bien être) ou de survie (Gouveia *et al*, 2014).

Les valeurs sont non seulement définies en théorie dans une perspective psychologique comme on l'a vu dans le paragraphe précédent, mais elles ont également été «mises en œuvre» et employées au moyen de l'éthique appliquée aux domaines professionnels. Par exemple, dans le domaine médical, qui utilise la bioéthique pour définir les valeurs importantes tenant lieu de principes directeurs pour les professionnels de la recherche biomédicale, comme les médecins, infirmiers / ères et professionnels de santé, Beauchamp & Walters (1999) définissent 4 valeurs comme base pour le cadre de la bioéthique: (1) *l'autonomie*: agir de manière intentionnée sans subir le contrôle d'influences qui agiraient à l'encontre d'un acte volontaire; (2) *la bienfaisance (valorisation)*: agir au profit de la société en général; (3) *la justice*: être équitable et raisonnable, et (4) *la non-maléficience (non-malfaisance)*: ne pas imposer intentionnellement de risques ou de torts à autrui.

En nous basant sur cette revue des publications nous avons choisi deux théories des valeurs pour notre étude; l'une tirée des recherches du domaine psychologique, et l'autre basée sur l'éthique appliquée, qui est une application pratique de la philosophie morale. La première théorie que nous avons choisie est celle de Cheng et Fleischmann (2010), car dans leur méta-inventaire des valeurs morales ils ont dressé une liste exhaustive de 16 valeurs humaines, qui est basée sur les valeurs trouvées dans 12 études distinctes. A notre avis, cette méta-analyse saisit les valeurs les plus importantes listées par d'autres chercheurs, et elle constitue un exemple empirique inspiré des publications psychologiques. La deuxième théorie des valeurs que

nous avons choisie est un exemple d'éthique appliquée qui a été largement pratiquée dans le domaine médical depuis plus de quarante ans. Nous voulons étudier son applicabilité aux armes autonomes, car les principes de bioéthique répondent à de nombreuses inquiétudes que les personnes peuvent avoir concernant les armes autonomes.

Tableau 2 Théories des valeurs (aperçu)

Author	Contribution majeure	Définition des valeurs	Valeurs
Schwartz (1994)	L'étude examine l'universalité potentielle des valeurs humaines et donne un ensemble de relations dynamiques entre ces valeurs. L'étude n'a pas trouvé d'aspect universel des valeurs, mais a trouvé ce qui peut justifier une quasi universalité de quatre valeurs d'un ordre supérieur. Également, l'étude a trouvé que tout montre que dix types de valeurs sont reconnus par de nombreuses personnes dans les sociétés contemporaines.	Les valeurs sont définies comme étant: *«des buts transituationnels à atteindre, d'importance variable, servant de principes directeurs dans la vie d'une personne ou d'une autre entité sociale»* (p. 21) Cinq caractéristiques contribuent à la définition des valeurs humaines: (1) *«Conviction (croyance»); (2) qui procède d'un état final recherché ou de modes de conduite, qui (3) transcende des situations spécifiques, (4) guide le choix ou l'évaluation du comportement, des personnes et des événements, et (5) est classée par importance par rapport à d'autres valeurs pour former un système de priorités dans les valeurs* (Schwartz, 1992; Schwartz & Bilski, 1987)»	1. *Pouvoir*: statut social et prestige, domination et contrôle des personnes et des ressources (autorité, richesse, pouvoir social).[4] 2. *Succès*. Succès personnel en prouvant ses compétences selon des critères sociaux (ambition, succès, capacité, influence). 3. *Hédonisme*: plaisir et satisfaction sensuelle au profit de soi (plaisir, belle vie, complaisance envers soi-même). 4. *Stimulation*: enthousiasme, nouveauté, et défis à relever (vie variée, vie pleine d'événements, audace). 5. *Indépendance*: pensée et actions indépendantes – choisir, créer, explorer (créativité, liberté, choisir ses objectifs, curieux, indépendant). 6. *Universalisme*: compréhension, appréciation, tolérance, et protection pour le bien-être de toutes les personnes et pour la nature (largesse d'esprit, justice sociale, égalité, paix dans le monde, monde de la beauté, harmonie avec la nature, sagesse, protection de l'environnement). 7. Bienveillance: préservation et amélioration du bien-être des personnes avec lesquelles on est en contact personnel fréquent (serviabilité, honnêteté, pardon, responsabilité, loyauté, amitié vraie, amour

[4] Les mots entre parenthèses sont les éléments de valeur spécifiques qui spécifient la valeur universelle.

		(p. 20).	mature). 8. *Tradition*: respect, engagement, acceptation des coutumes et des idées que les cultures et religions traditionnelles portent en elles (respect de la tradition, humilité, dévotion, acceptation de son propre rôle dans la vie). 9. *Respect*: retenue dans ses actes, ses inclinations et ses impulsions qui sont susceptibles de choquer ou nuire à autrui et de violer les codes de conduite or les normes (obéissance, autodiscipline, politesse, respect des parents et des aînés). 10. *Sécurité*: sûreté, harmonie et stabilité de la société, dans les relations et pour soi-même (ordre social, sécurité familiale, sécurité nationale, réciprocité des services rendus).
Friedman, Kahn Jr, Borning, et Huldtgren (2013)	Un aperçu de l'approche de la conception éthique (VSD), et indications pour une application pratique. Fournit des informations pour permettre à d'autres chercheurs d'utiliser et d'élargir la VSD, et aux acteurs concernés de prendre les valeurs en compte dans le traitement de l'information et la conception de systèmes in-	Une valeur est définie au sens large en ce qu'elle *«désigne ce qu'une personne ou un groupe de personnes considèrent comme important dans la vie»* (p. 57). La méthode VSD concerne en particulier les valeurs morales qui sont: *«des questions ayant trait à l'équité, à la justice, au bien-être des personnes et à la vertu, englobant dans la théorie philoso-*	1. Bien-être des personnes: se rapporte aux bonnes conditions de vie physique, matérielle et psychologique des personnes. 2. Propriété: droit de posséder un objet (ou de l'information), de l'utiliser, d'en tirer un revenu, et de le léguer. 3. Intimité: autorisation, droit qu'a une personne de décider de quelles informations le ou la concernant peuvent être communiques à autrui. 4. Liberté vis-à-vis de toute partialité, c'est-à-dire de toute injustice perpétrée contre des individus ou des groupes, y compris de par des préjugés sociaux, techniques ou sociaux nouvellement apparus.

	formatiques.	*phique de la morale la déontologie, le conséquentialisme et la vertu».* (p.72).	5. Possibilité d'accès universelle: faire en sorte que tout le monde puisse utiliser correctement l'informatique. 6. Confiance: ce que chacun attend des personnes qui peuvent faire preuve de bonne volonté en l'élargissant vers les autres, se sentir vulnérables ou être victime de trahison. 7. Autonomie: capacité des personnes à décider, planifier et agir de manière à penser que cela va les aider à atteindre leurs objectifs. 8. Consentement éclairé: recueillir l'accord des personnes; ceci englobe les critères de divulgation et de compréhension (pour «éclairé»), et d'acte volontaire, de compétence et d'accord (pour «consentement»). 9. Imputabilité: les mesures qui font en sorte que les actions d'une personne, de plusieurs personnes ou d'une institution peuvent être ramenées uniquement à cette / ces personne(s) ou à cette institution. 10. Courtoisie: se conduire envers autrui avec politesse et considération. 11. Identité. La conscience que chacun a de qui il est dans la durée, de manière continue ou discontinue dans le temps. 12. Quiétude. Etat psychologique paisible et serein. 13. Durabilité environnementale: protéger les écosystèmes pour qu'ils répondent aux besoins présents sans compromettre les générations.

Graham *et al.* (2012)	Présentation (avec les critiques et résultat empirique) de la **théorie du fondement de la morale.** Cette théorie peut être utilisée pour connaître le jugement moral des personnes. Elle est décrite comme étant une approche de la morale pluraliste, nativiste, culturo-«développementaliste» et intuitionniste.	Pour représenter les cinq concepts de cette théorie du fondement moral, le terme «fondement» («foundation») est choisi, mais il est interchangeable avec «valeur» ou «vertu». Aucune définition précise de «fondement» n'est donnée, mais le terme est utilisé comme métaphore architecturale («fondation») pour affirmer que «*la théorie du fondement moral s'intéresse à la première esquisse universelle de l'esprit moral, et à la manière dont cette esquisse est révisée à des degrés variables selon les cultures*» (p. 10).	Les cinq fondements de la théorie sont les suivants: 1. Fondement bienveillance / malveillance: capacité à ressentir la douleur des autres personnes; sous-tend les vertus de bonté, d'amabilité et dévouement. 2. Fondement équité / tromperie: principe de réciprocité altruiste, génère les idées de justice, de droits et d'autonomie. 3. Fondement loyauté / trahison: renvoie aux coalitions changeantes, sous-tend les vertus de patriotisme et de sacrifice de soi au profit du groupe. 4. Fondement autorité / subversion: renvoie aux interactions sociales de hiérarchie, et sous-tend les vertus d'exercice de l'autorité et de camaraderie, y compris le respect de l'autorité légitime et des traditions. 5. Fondement respect du sacré / dégradation: renvoie à la psychologie de l'aversion et de la contamination; sous-tend les concepts religieux prônant de plus hautes aspirations, d'une vie plus détachée de la chair, us détachée de la chair, plus «noble».
Cheng et Fleischmann (2010)	Méta-analyse de 12 inventaires des valeurs humaines. Cette étude propose un méta-inventaire des valeurs humaines.	Donne un résumé des définitions des valeurs: *«les valeurs servent de principes directeurs pour ce que les personnes considèrent comme étant important dans la vie»* (p.2).	(1) Liberté, (2) serviabilité, (3) réussite, (4) honnêteté, (5) respect de soi, (6) intelligence, (7) ouverture d'esprit, (8) créativité, (9) égalité, (10) responsabilité, (11) ordre social, (12) richesse, (13) compétence, (14) justice, (15) sécurité, et (16) spiritualité.

Gouveia, Milfont, and Guerra (2014)	Etude empirique des valeurs fondamentales. L'étude propose un cadre «trois par deux» qui contient six sous-catégories de valeurs fondamentales.	Deux fonctions primaires des valeurs sont identifiées: (1) elles guident les actes, et (2) elles sont des expressions cognitives des besoins.	Objectifs personnels – valeurs de besoins de prospérité: émotion, plaisir, sexualité. Objectifs personnels – valeurs de besoins pour la survie: pouvoir, prestige, succès. Objectifs centraux – valeurs de besoins de prospérité: beauté, connaissance, maturité. Objectifs centraux: valeurs de besoins pour la survie: santé, stabilité, survie. Objectifs sociaux: valeurs de besoin de prospérité: affectivité, appartenance, soutien. Objectifs sociaux: valeurs de besoins pour la survie: obéissance, religiosité, tradition.
Beauchamp et Walters (1999)	Cet article constitue le premier chapitre d'un livre sur la bioéthique. Ce chapitre décrit trois principes moraux qui donnent un cadre pouvant être utilisé pour réfléchir sur les problèmes de bioéthique.	Les auteurs utilisent les termes «principes» et «valeurs» comme synonymes. Ils définissent un principe comme «*un principe est un modèle fondamental de conduite sur lequel un grand nombre d'autres modèles et jugements moraux s'appuient pour leur défense et validité*» p. 17.	1. *Autonomie: agir intentionnellement sans interférences influentes qui agiraient à l'encontre d'un acte volontaire.* 2. *Bénéficence / bienfaisance: agir pour le bien de la société prise dans son ensemble.* 3. *Justice: être équitable et raisonnable.* 4. *Non-malfaisance: ne pas imposer intentionnellement des risques ou des torts à autrui.*

2.2.3. Valeurs associées aux armes autonomes

Les valeurs telles qu'elles sont définies dans les théories des valeurs dans la section 2.2.2 ne sont pas souvent évoquées explicitement dans les publications sur les armes autonomes comme le montre l'aperçu donné dans le Tableau 3, mais la plupart des études discutent des différentes valeurs et les problèmes éthiques qui s'y rapportent. Deux rapports publics émanant de Human Rights Watch mentionnent l'absence d'*émotion humaine, d'imputabilité, de responsabilité, le manque de dignité humaine et le mal causé aux personnes* comme valeurs mises en jeu par les armes autonomes (Docherty, 2012, 2015). Sharkey et Suchman (2013) affirment que les valeurs d'imputabilité (accountability) et de responsabilité (responsibility) sont importantes lors de la conception de systèmes robotisés pour les opérations militaires.

Dans le domaine de l'éthique militaire, Johnson et Axinn (2013) donnent la *responsabilité, l'atténuation du mal causé à autrui, la dignité humaine, l'honneur et le sacrifice humain* comme étant des valeurs dans leur discussion sur le fait de savoir si la décision de tuer un être humain doit être confiée à une machine ou non. Cummings (2006), dans son étude de cas concernant le missile tactique Tomahawk, se réfère aux valeurs universelles proposées par Friedman et Kahn Jr. (2003) et écrit qu'après *l'imputabilité et le consentement éclairé*, la valeur de *bien-être humain* est la valeur fondamentale pour les ingénieurs qui développent des armes, vu qu'elle se rapporte à la santé, à la sécurité et au bien-être du public. Elle avance également que les principes juridiques de *proportionnalité et de discernement* sont à considérer dans le contexte de la conception d'armes. La *proportionnalité* renvoie au fait qu'une attaque est justifiée que lorsque les dommages ne sont pas jugés excessifs. Le *discernement* («discrimination») signifie qu'une distinction entre combattants et non-combattants est possible (Hurka, 2005). Asaro (2012) se réfère également aux principes de *proportionnalité* et de *discerne-*

ment, et affirme que les armes autonomes ouvrent un espace moral dans lequel de nouvelles normes sont nécessaires. Bien qu'il n'évoque pas explicitement les valeurs dans son argumentation, il se réfère à la valeur de *la vie humaine* et à la nécessité qu'il y a à ce que les êtres humains soient partie prenante dans la décision de supprimer une vie humaine. D'autres études décrivent surtout des problèmes éthiques, comme la *prévention du tort fait à autrui* («harm prevention»), la *préservation de la dignité humaine*, *la sécurité*, la *valeur de la vie humaine* et l'*imputabilité* (Horowitz, 2016; UNDIR, 2015; James Walsh & Schulzke, 2015; Williams, Scharre & Mayer, 2015).

En se référant à la revue des publications des valeurs évoquées dans les publications existantes et mises en jeu par les armes autonomes, les valeurs dignité humaine, tort fait à autrui, sécurité, responsabilité et imputabilité seront ajoutées à la liste des valeurs qui seront utilisées dans cette étude, car ces valeurs sont évoquées plus d'une fois dans les données du Tableau 3. Avec les valeurs provenant du domaine de la bioéthique et du méta-inventaire de Cheng et Fleischmann (2010), ces valeurs constitueront notre point de départ dans l'investigation empirique de cette étude.

Tableau 3 – Valeurs mises en jeu par les armes

Author	Contribution majeure	Définition des valeurs	Valeurs
Cummings (2006)	Application de l'approche VSD au problème de la conception du missile tactique Tomahawk. L'étude expose les considérations touchant aux problèmes éthiques dans le processus de conception pour les instructeurs et les opérateurs.	N/ a	Dans la liste de Friedmann et al (2006), les valeurs qui s'appliquent à la conception de systèmes d'armes sont l'imputabilité, le consentement éclairé, mais la plus importante est le bien-être humain. Les principes de discernement et de proportionnalité sont importants dans la considération «bien-être humain».
Docherty (2012)	Rapport émanant de Human Rights Watch dans lequel des aspects du droit humanitaire international et des problèmes éthiques dérivés des armes autonomes sont évoqués.	Aucune définition du terme «valeur», mais le texte évoque les valeurs et les questions éthiques.	Valeurs / questions éthiques : • Absence d'émotions humaines • Imputabilité • Responsabilité.
Johnson et Axinn (2013)	Etude sur les questions éthiques soulevées par l'usage des armes autonomes létales robotisées. Aborde la question de savoir si la décision de supprimer une vie humaine peut être confiée à des machines.	Aucune définition du terme «valeur», mais le texte évoque les valeurs.	Valeurs / questions éthiques • Responsabilité • Atténuation du tort causé • Dignité humaine • Honneur • Sacrifice humain

Sharkey et Suchman (2013)	Etude sur la définition et la conception de l'autonomie et de l'imputabilité dans les systèmes robotisés pour les opérations militaires.	Aucune définition du terme «valeur», mais le texte évoque les valeurs et les questions éthiques.	Valeurs: • Imputabilité • Responsabilité
Docherty (2015)	Rapport de Human Rights Watch dans lequel sont évoqués l'absence d'imputabilité et problèmes éthiques pour les armes autonomes.	Aucune définition du terme «valeur», mais le texte évoque les valeurs et les questions éthiques.	Valeurs / questions éthiques : • Manque de dignité humaine • Imputabilité • Responsabilité • Tort causé
United Nations Institute for Disarmament Research (2015)	L'étude met en lumière quelques questions éthiques et sociales concernant la militarisation des technologies autonomes. La réflexion éthique sur les valeurs culturelles et sociales de la militarisation des technologies autonomes est encouragée.	Aucune définition du terme «valeur», mais le texte n'évoque pas explicitement les valeurs, mais quelques questions éthiques sont évoquées.	Les questions éthiques qui sont évoquées sont : • Atténuation ou élimination des torts causés • Considération pour la conscience publique • Offense à la dignité humaine (lorsque la suppression d'une vie se fait sans intention humaine).
Walsh et Schulzke (2015)	Etude expérimentale pour savoir si les civils américains sont davantage susceptibles de se lancer dans une guerre lorsque des drones sont utilisés. Etude empirique de grande ampleur qui aborde l'éthique des frappes par drones.	Aucune définition du terme «valeur», mais le texte n'évoque pas explicitement les valeurs, mais quelques questions éthiques sont évoquées.	Questions d'ordre éthique : • Sécurité • Respect de l'immunité des civils • Prévention des torts

Williams, Scharre, et Mayer (2015)	Discute plusieurs questions éthiques concernant le développement des systèmes autonomes, et donne des recommandations à l'intention des décideurs de la politique de défense.	Aucune définition du terme «valeurs», mais le texte évoque plusieurs valeurs qui sont utilisées de manière interchangeable avec les questions éthiques.	Valeurs / questions éthiques • Sécurité • Tort causé • Valeur de la vie humaine (les personnes ont le droit d'être tuées par d'autres personnes.
Horowitz (2016)	Description du débat sur les implications éthiques des armes autonomes. Considère les systèmes d'armes létaux autonomes (LAWS) sous trois catégories: munitions, plates-formes et systèmes opérationnels. Ceci clarifie le débat et décrit deux questions éthiques.	Aucune définition du terme «valeur», le texte n'évoque pas explicitement les valeurs, mais quelques questions éthiques sont évoquées.	Valeurs : • Imputabilité (les systèmes autonomes ne sont pas contrôlés par l'homme de manière significative, par conséquent ils créent un vide d'imputabilité) • Dignité humaine (les gens ont le droit d'être tués par une personne qui a décidé de les tuer).

2.2.4. Hiérarchie des valeurs

Une des approches pour déterminer quelles sont les valeurs utiles pour la conception des armes autonomes est la translation des valeurs en exigences dans la conception, qui peuvent être rendues visibles grâce à une hiérarchie des valeurs (Van de Poel, 2013). Cette structure hiérarchique des valeurs, des normes et des exigences dans la conception rend les jugements sur les valeurs nécessaires à cette translation explicites, transparents et discutables. Pour ce faire, les valeurs qui sont définies dans le langage habituel devront être traduites en «valeurs formelles en langage formel» (Aldewereld, Dignum &Tan, 2015, p. 835). Une manière de formaliser les valeurs en normes serait une convention de règles représentées comme: «X compte pour Y», ou «X compte pour Y dans le contexte C» (Searle, 1995, p. 28). Le fait d'expliciter les valeurs en règles formelles permet une réflexion critique dans les débats et fait ressortir les jugements de valeurs sur lesquels portent les désaccords. La transparence est importante, comme le souligne Van de Poel (2013, p. 265) avec éloquence: «bien que des choix transparents ne sont pas forcément meilleurs ou plus acceptables, la transparence semble être une condition minimale dans une société démocratique qui essaie de protéger ou d'améliorer l'autonomie morale des citoyens, surtout dans les cas où la conception affecte la vie des autres en plus de celle des concepteurs, et cela arrive souvent.»

Le niveau supérieur de la hiérarchie des valeurs consiste en les valeurs, comme il apparaît sur la Figure 3, le niveau intermédiaire contient les normes, qui peuvent être des capacités, des propriétés ou attributs de l'artefact, et le niveau inférieur fait figurer les exigences dans la conception qui peuvent être identifiées. La relation entre les niveaux n'est pas de nature déductive et peut être figurée de manière descendante, en ajoutant des instructions, ou ascendante en cherchant les motiva-

tions et les justifications des exigences du niveau inférieur. La conceptualisation ascendante est une activité philosophique qui ne nécessite pas de connaissances spécialisées, et la spécification descendante des valeurs exige des connaissances du contexte ou du domaine spécifique pour ajouter du contenu à la conception. (Van de Poel, 201.)

Van de Poel (2013, p.262) définit la spécification comme étant: «la translation d'une valeur générale en l'une ou plusieurs des nécessités spécifiques à la conception», et indique que cela peut se faire en deux phases:

1. Transposer une *valeur générale* en une ou plusieurs *normes générales*;
2. Transposer ces *normes générales* en plusieurs *nécessités spécifiques à la conception.*

Deux critères sont pertinents pour la phase 1: (1) la norme doit être une réponse appropriée à la valeur, et (2) la norme doit être une réponse suffisante à la valeur. Dans la phase 2, l'exigence doit être plus spécifique en ce qui concerne le champ d'applicabilité, les objectifs recherchés et les actions à accomplir pour atteindre les buts de la norme (Van de Poel, 2013).

Cette translation peut s'avérer difficile, comme une connaissance est nécessaire dans l'utilisation et le contexte prévu de la valeur, qui n'est pas toujours évident au début d'un projet de conception. Aussi bien, comme les artefacts sont souvent utilisé d'une manière ou dans un contexte non prévus, de nouvelles valeurs apparaissent, ou une absence de valeurs est constatée (Van Wynsberghe & Robbins, 2014). Un exemple en est les drones qui ont été initialement conçus pour une utilisation militaire, mais qui sont également utilisés par des civils pour filmer des événements, et même comme éclairage de fond pour le spectacle de mi-temps du Super Bowl 2017. La valeur

sécurité est interprétée différemment par les utilisateurs militaires qui les emploient dans des régions désertes, comparée à celle qui concerne les 300 drones survolant en formation un stade de football dans une zone peuplée. Les différents contextes et utilisations d'un drone conduiront à une interprétation différente du mot valeur, et pourraient conduire à des normes de distance plus strictes pour la sécurité des vols; ceci pourrait en conséquence être plus précisément spécifié pour les exigences de conception alternative pour les rotors et les logiciels pour les alertes de proximité, pour ne citer que deux exemples.

L'application d'une hiérarchie des valeurs aux armes autonomes est illustrée par un exemple dans lequel la valeur d'imputabilité est transposée en normes «pour la transparence des prises de décision» et «connaissance de l'algorithme». Cette translation permettra aux utilisateurs de comprendre les choix dans les décisions que fait l'arme autonome pour préciser et justifier ses actions (Figure 3). Les normes pour la transparence de la prise de décision conduisent à des exigences de conception spécifiques. Dans ce cas, il faudra visualiser l'arbre décisionnel, mais aussi présenter les variables de décision que l'arme autonome a utilisées, comme les compromis concernant les pourcentages de dégâts collatéraux des différents scénarios d'attaque, pour renseigner sur la proportionnalité d'une attaque. L'arme autonome devrait également être capable de présenter les informations fournies par les capteurs, par exemple l'imagerie du site, pour montrer qu'elle a discerné les combattants des non-combattants. Pour renseigner sur l'algorithme, une arme autonome devrait être conçue avec des caractéristiques dont elle n'est pas dotée normalement. Dans ce cas ces caractéristiques doivent inclure un écran comme interface pour l'utilisateur qui permet de visualiser l'algorithme sous forme lisible par les hommes, et la fonctionnalité de téléchargement des changements apportés par l'algorithme comme faisant par-

tie de ses capacités d'apprentissage en tant que machine, qui peuvent être étudiées par un tiers indépendant, comme un tribunal pénal des Nations Unies, si la légalité des actions d'une arme autonome est mise en question.

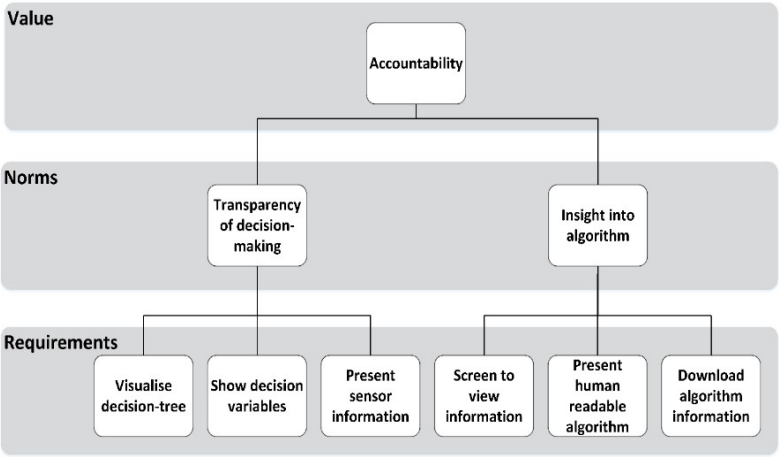


Figure 3 – Hiérarchie des valeurs pour la valeur de «responsabilité» dans la conception d'armes autonomes

Kroes et Van de Poel (2015) affirment qu'une mesure objective des valeurs n'est pas possible du fait que l'opérationnalisation est effectuée au moyen de jugements de valeurs de second ordre, ce qui affaiblit sérieusement la validité du concept de mesure des valeurs. Les jugements sont souvent considérés comme subjectifs du fait que leur vérité, tout comme leur fausseté, dépend des sentiments et des attitudes de la personne qui juge (Searle, 1995). Pour compenser ce manque de validité, le concepteur pourrait se tourner vers des codes et des modèles techniques élaborés par des commissions et qui représentent des références raisonnables pour l'opérationnalisation et la me-

sure des valeurs dans la conception. Cependant, les modèles peuvent ne pas refléter les développements techniques et sociaux les plus récents, et l'opérationnalisation nécessite des jugements de valeurs de la part des concepteurs. Kroes et Van de Poel (2015, p. 177) conseillent de les «*intégrer dans un réseau d'autres considérations, y compris les définitions des valeurs mises en jeu dans la philosophie morale (ou les lois), dans les codes et les modèles existants, les expériences de conception qui ont précédé, etc.*»

Dans la présente étude, nous suivons le conseil de Kroes et Van de Poel (2015), et n'appliquons pas strictement la hiérarchie des valeurs comme méthode pour préciser, concevoir et tester les exigences pour les armes autonomes, mais nous utilisons l'exemple de la figure 3 pour établir le lien entre les phases d'investigation empirique et conceptuelle de l'approche de la conception éthique de notre recherche. La hiérarchie des valeurs nous permet de voir au-delà et de transposer le concept abstrait des valeurs étudiées dans la phase d'investigation conceptuelle et d'en faire un exemple concret d'arme autonome que nous pouvons utiliser pour contribuer à la phase d'investigation empirique de notre recherche. La hiérarchie de valeurs de la Figure 3 est utilisée pour faire office d'orientation, d'inspiration et d'indicateur de direction pour notre recherche.

2.3. L'agentivité

Le concept d'«agentivité» est étudié dans de nombreux domines académiques; dans tous les cas l'agentivité est étudiée sous des angles différents. Nous avons passé en revue les publications dans les domaines de la psychologie cognitive, de l'intelligence artificielle et de la philosophie morale. Nous avons résumé les caractéristiques de l'agentivité rencontrées dans ces trois domaines dans le Tableau 4, et en conclusion nous dirons pourquoi nous avons choisi ces caractéristiques.

2.3.1. L'agentivité en psychologie cognitive

Dans le domaine de la psychologie cognitive, les dimensions de la perception mentale (intellectuelle) ont été étudiées pour la première fois par H.M. Gray *et al* (2007). Ils constatèrent que la perception mentale a deux dimensions: (1) L*'expérience*, incluant les facteurs faim, peur, douleur, plaisir, colère, désir, personnalité, conscience, orgueil, gêne et joie, et (2) L*'agentivité*, qui englobe le contrôle de soi, la morale, la mémoire, l'émotion, la reconnaissance des choses, la prévision, la communication et la pensée. Alors que l'expérience est qualifiée de passivité («patiency») et associée aux droits et privilèges, l'agentivité est liée à la capacité d'agir (sur le plan moral), et associée à la responsabilité. Plusieurs études en psychologie cognitive montrent que les personnes attribuent l'agentivité à des non-humains, et perçoivent ces «agents» non-humains comme ayant les mêmes capacités d'agentivité que les humains face à des dilemmes d'ordre moral (Hristova & Grinberg, 2015). Ceci ne se limite pas seulement aux êtres vivants comme les animaux, mais l'agentivité est également attribuée aux objets inorganiques comme les robots et les zombies (K. Gray & Wegner 2012; Waytz *et al*, 2010).

2.3.2. L'agentivité en matière d'Intelligence Artificielle

Les chercheurs dans le domaine de l'intelligence artificielle ont également étudié les caractéristiques des agents, mais davantage d'un point de vue informatique, pour créer l'architecture d'un agent rationnel capable de raisonner et de réfléchir face à une alternative. Bratman, Israel et Pollack (1998) décrivent une architecture Conviction – Désir – Intention pour le processus de réflexion à l'intention d'agents limités dans le traitement de l'information et les délais de réaction. *«Les convictions sont le statut des propriétés de son propre monde (et de lui-même) que l'agent croit être vraies»* (Dignuù, Kinny, & Sonenberg, 2002, p. 407). Perugini et

Bagozzi (2004, p. 71) définissent un désir comme *«un état d'esprit par lequel un agent ressent une motivation personnelle pour exécuter une action ou atteindre un but.»* Les intentions impliquent un engagement à un plan, et impliquent une forme de «planification» pour atteindre ce but (Bratman *et al*, 1988; Perugini & Bagozzi, 2004); Dignum *et al*, (2002) élargissent le cadre Conviction – Désir – Intention pour y incorporer des concepts sociaux tels que des normes et obligations, car ce sont des concepts important pour relier les agents dans un système multi-agents.

2.3.3. L'agentivité en philosophie morale

La philosophie morale est le troisième domaine d'étude qui a défini les propriétés des agents. Sullins (2006) pose trois prérequis pour déterminer si un robot peut être considéré comme agent moral: (1) *Autonomie*: le robot est significativement indépendant de ses programmeurs, opérateurs et utilisateurs; (2) *Intentionnalité*: la programmation «faire le bien ou du tort»; (3) *Responsabilité*: joue un rôle social et est responsable d'autres agents moraux. Himma (2009) définit l'agentivité comme *«être capable de faire quelque chose qui compte comme un acte ou une action»*. La notion d'agentivité morale est que l'on est susceptible de rendre compte de son comportement qui est gouverné par un code moral. Deux capacités sont nécessaires à l'agentivité morale. La première est de choisir librement vos actions, résultats d'une réflexion. La seconde capacité est de comprendre les concepts moraux, comme la distinction entre le «bien» et le «mal», ou le «bon» et le «mauvais», mais aussi d'appliquer des principes moraux tels que «il est mal de faire du tort de manière intentionnée» (Himma, 2009).

Auteur	Caractéristiques Agentivité	Champ scientifique
Bratman *et al.* (1988)	Convictions, désirs, intentions	Intelligence artificielle
Bandura (2001)	Intentionnalité, prévoyance, autorégulation, réflexion sur ses propres actes	Psychologie cognitive
F. Dignum *et al.* (2002)	Croyances, désirs, intentions, normes	Intelligence artificielle
Sullins (2006)	Autonomie, intentions, responsabilité	Philosophie morale
H. M. Gray *et al.* (2007)	Contrôle de soi, moralité, mémoire, émotion, reconnaissance des choses, planification, communication, pensée.	Psychologie cognitive
Himma (2009)	Libres choix, délibération, compréhension, application des règles morales	Philosophie morale
Waytz *et al.* (2010)	Capacité de planifier et d'agir.	Psychologie cognitive

Tableau 4 – Présentation des caractéristiques agentivité

Les principales caractéristiques de l'agentivité ressortant de l'analyse des publications sur la perception de l'agentivité dans les domaines de la psychologie cognitive, de l'intelligence artificielle et de la philosophie morale sont classées dans l'ordre chronologique dans le Tableau 4. Nous avons sélectionné tous les concepts utilisés par les auteurs pour définir ou caractériser

l'agentivité, et nous ne les avons pas filtrés à cette phase de l'étude. Toutes ces caractéristiques seront utilisées dans la phase empirique pour étudier la perception de l'agentivité des armes autonomes.

3. Méthode

> *«Toutes les questions sont au fond des variations sur une seule question d'ordre éthique: un drone peut-il de manière autonome décider de «faire feu» sans intervention humaine? A mon avis, la réponse est «non». Il devrait toujours y avoir une interface humaine. La guerre et les conflits ne sont pas un jeu sur ordinateur, la guerre est une interaction sociale entre êtres humains. Un drone n'est qu'un simple instrument, un outil qui ne peut jamais opérer de manière autonome».*
>
> *Personne sondée, étude finale*

Cette section présente la méthodologie, les hypothèses, le concept de recherche, l'opérationnalisation des scénarios et des concepts, l'approche analytique, l'enregistrement de l'étude et conclut sur les questions méthodologiques.

3.1. Méthodologie

Les méthodes utilisées dans les diverses parties de la présente étude sont décrites dans cette sous-section, en commençant par la revue des publications existantes, puis l'enquête en ligne sur les valeurs, les entretiens d'experts, le codage des entretiens, et enfin la méthode basée sur les expériences contrôlées randomisées.

3.1.1. Revue des publications

Google Scholar a été utilisé pour chercher des articles pour la revue des publications; les premiers mots-clés entrés *Value-sensitive* (= éthique, qui tient compte des valeurs) ET *Design* (conception), ce qui a fourni quelques articles sur la conception éthique. Le *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (Manuel de

l'éthique, des valeurs et de la conception technologique: sources, théorie, valeurs et domaines d'application) a été utilisé pour étoffer les connaissances sur le sujet et trouver des références supplémentaires. Ensuite, nous avons cherché des articles sur les valeurs humaines fondamentales avec les mots-clés *human* ET *values*, et *universal* ET *values*, mais cela n'a pas donné beaucoup de résultats. Nous avons analysé quelques théories connues sur les valeurs humaines fondamentales et les hiérarchies des valeurs. En prenant ces articles, nous avons utilisé l'option «cité par» dans Google Scholar pour trouver des articles qui utilisaient ceux-ci comme sources. Cela a donné des résultats intéressants sur des articles définissant les valeurs humaines de base, et nous avons choisi les articles qui faisaient apparaître une liste des valeurs. Les articles qui ne faisaient que reprendre des études précédentes, par exemple sur la théorie de Schwartz, n'ont pas été retenus. Ensuite, nous avons cherché des articles concernant les valeurs des personnes relativement aux armes en choisissant les mots clés *value* ET *weapon* (arme), et les mots clés *Value-sensitive* ET *Weapons*. Nous avons retenu les articles qui contenaient le mot-clé *value*.

3.1.2. Enquête «en ligne» sur les valeurs

L'enquête en ligne a été effectuée en utilisant l'outil d'enquête en ligne Qualtrics[5] fournit par le MIT. Après quatre questions sur des données démographiques, nous avons posé aux personnes interrogées trois questions sur les valeurs qu'ils associaient aux armes autonomes. La première question était de classer les quatre valeurs de bioéthique *Autonomie, Non-malfaisance, Bienfaisance et Justice* de la plus applicable aux armes autonomes à la moins applicable. Dans la deuxième question nous avons demandé de choisir les cinq valeurs qui

[5] https://www.qualtrics.com/

s'appliquent le mieux aux armes autonomes d'après la liste de Cheng et Fleischmann (2010). La troisième question était une question ouverte pour laquelle nous avons demandé de donner une autre valeur qui ne figurait pas dans les deux premières questions. Cette étude a été vérifiée deux fois. En premier lieu, deux camarades d'étude ont revu les questions, et après le traitement de leurs feedback, l'enquête a été vérifiée une seconde fois par différentes personnes, plus représentatives de futurs participants. Le questionnaire a été distribué via les réseaux sociaux, comme Facebook, Linkedin et Twitter, il a été envoyé à notre blog personnel et la plateforme en ligne «Call for Participants» a été utilisée pour la distribution en dehors de mon réseau social pour récupérer plus de participants. Au cours de la semaine organisées par les membres du Media Lab du MIT en avril 2017, les personnes intéressées par notre recherche ont été priées de participer à notre enquête, et nous avons également envoyé le lien de notre enquête au Groupe de Coopération Evolutive «Slack».[6] L'enquête en ligne s'est déroulée pendant 3 semaines et demie, du 17 mars au 11 avril 2017.

3.1.3. Entretiens avec spécialistes

Nous avons mené six entretiens semi-structurés avec des spécialistes, en suivant la méthode *Photo Elicitation Interview* (PEI) (Harper, 2002) pour recevoir des opinions différenciées d'experts dans le domaine des armes autonomes, et avoir une compréhension plus profonde des problèmes de leurs points de vue. La méthode PEI peut être utilisée pour accéder à une compréhension partagée des valeurs, et pour mener une discussion entre le concepteur et les parties prenantes sur ces valeurs (Pommeranz, Detweiler, Wiggers & Jonker, 2011). Ceci

[6] «Slack» est un outil de collaboration fonctionnant sur le cloud, qui porte la communication en équipes (https://slack.com/).

donne plus de force à ce que disent les parties prenantes, et tempère ce qu'avancent les chercheurs (Le Dantec, Poole & Wyche, 2009). Nous avons demandé aux experts de choisir trois photos avant l'entretien, et au cas où ils n'en choisiraient aucune eux-mêmes nous avons aussi choisi huit photos d'armes autonomes. Durant les entretiens, nous avons saisi les valeurs des experts concernant les armes autonomes en leur demandant pourquoi ils avaient choisi telle ou telle photo, et quelle valeur elle représentait pour eux.

Tous les six ont fait l'effort de choisir trois photos avant l'entretien. Durant les entretiens, nous avons remarqué qu'ils avaient beaucoup réfléchi avant de choisir. En discutant sur les photos, nous avons abordé un grand nombre de questions diverses, comme l'intuition, les scénarios futuristes, l'économie, la tactique militaire, les films de science-fiction et même la politique. Même si toutes les photos ne conduisaient pas à des discussions sur les valeurs humaines fondamentales telles qu'on les trouve évoquées dans les publications, la méthode PEI a bien fait ressortir des informations sur les principes que les spécialistes considèrent comme importants concernant les armes autonomes. Elle a également permis une compréhension réciproque et un bon climat durant les entretiens.

3.1.4. Processus de codage

Les résultats des entretiens ont été traités au moyen de la méthode Values Coding décrite par Saldana (2015). Tenir un mémo (Appendice G. Mémos de codage) dans lequel le processus et les hypothèses sont listés est une partie intégrante d'une méthode de codage. Nous sommes partis d'une liste de codes préétablis qui contenait: «*(1) une valeur: l'importance que l'on attribue à soi-même, à une autre personne, chose ou idée, (2) attitude: la manière dont nous pensons et nous nous sentons vis-à-vis de nous-mêmes, d'une*

autre personne, chose ou idée, et (3) une conviction (croyance): cette partie du système qui englobe nos valeurs et attitudes plus nos connaissances, expériences, opinions, préjugés personnels, et les autres perceptions interprétatrices du monde social» (Saldana, 2015, p.89). Durant le codage des entretiens, il faut que les codes s'adaptent aux données, et non l'inverse. Nous avons donc ajouté un code «définition» à la liste des codes qui apparaissait. Nous avons pris des notes en codifiant les entretiens qui décrivaient nos démarches et hypothèses. Nous avons demandé à un deuxième chercheur, étudiant lui aussi en master TPM connaissant les valeurs, d'encoder les entretiens en utilisant la méthode «Value Coding» et aussi de tenir un mémo (Appendice G, Mémos d'encodage). Nous avons comparé les valeurs que nous avions encodées à celles du deuxième chercheur, et avons tiré les valeurs qui étaient similaires (Appendice H. Résultats du processus d'encodage).

3.1.5. Expériences contrôlées randomisées

La méthode utilisée pour les études pilotes et l'étude finale est appelée «expérience contrôlée randomisée». Oehlert (2010) donne quatre raisons pour créer des expériences: (1) elles permettent des comparaisons directes entre les traitements intéressants, (2) elles peuvent être conçues pour minimiser toute distorsion dans les comparaisons, (3) elles peuvent être conçues pour réduire l'erreur dans la comparaison, et (4) l'expérience est sous notre contrôle, ce qui nous permet de faire des déductions plus solides sur la nature des différences que l'on observe, en particulier sur la causalité. Ce dernier point permet de distinguer une expérience d'une étude basée sur l'observation. Les différentes procédures que nous visons à comparer sont dans ce sens un processus de traitement. Il est important que les effets d'un traitement puissent être attribués uniquement à une cause qui peut être mesurée, et aussi que les effets ne puissent pas être attribués à des causes multiples. Ceci est appelé facteur

de confusion que Oelhert (2010, p. 14) définit comme *«survenant lorsque l'effet d'un facteur ou d'un traitement ne peut pas être distingué de celui d'un autre facteur ou traitement»*. La randomisation contribue à s'assurer qu'un scénario soit attribué aux participants par hasard, et non sur la base de caractéristiques préétablies, comme le moment ou le lieu où l'enquête est menée. Nous utilisons la randomisation dans les études pour varier l'ordre des scénarios et l'ordre des questions posées aux personnes interrogées au moyen d'un dispositif basé sur les probabilités. Dans les études, nous avons choisi de distribuer un nombre égal de scénarios aux personnes interrogées.

3.2. Hypothèses

Du fait de la nature exploratoire de cette étude, nous avons développé une hypothèse pour la perception de l'agentivité de l'étude finale seulement, et non pour les études pilotes ou les variables dépendantes. Les études pilotes ont été utilisées pour établir quelles manipulations auraient quels effets, et pour vérifier si la formulation des scénarios ainsi que les questions étaient clairs et efficaces.

En avançant que les personnels militaires verraient les armes autonomes comme n'importe quelle autre arme, et donc comme rien de plus qu'un outil pour obtenir un effet, nous émettons l'hypothèse dans l'étude finale que les personnels militaires ne percevront pas les armes autonomes comme ayant des états mentaux. Par conséquent on peut s'attendre à ce qu'il n'y ait pas de différence entre la condition d'une arme autonome à agentivité neutre et la condition dans laquelle un être humain actionne un drone à distance. On peut également s'attendre à ce que la condition d'agentivité neutre soit jugée comme étant significativement différente de la condition d'agentivité élevée, dans laquelle nous disons précisément aux participants que l'arme autonome a des caractéristiques

d'agentivité, comme la capacité à planifier et fixer ses propres objectifs. A partir de ce raisonnement, nous avons émis l'hypothèse que:

> *H1: les personnels militaires ne percevront pas les armes autonomes comme ayant des états mentaux.*

3.3. Conception de recherche

Nous avons testé plusieurs conceptions de recherche dans les deux études pilotes avant de décider de la conception de recherche pour l'étude finale qui est décrite Figure 3. Pour être succinct, nous ne faisons pas figurer les deux conceptions pour les études pilotes dans cette section, et nous faisons figurer uniquement le modèle de recherche concernant le niveau élevé. Le modèle de recherche de l'étude finale inclut la perception de l'agentivité en tant que variable indépendante, les valeurs morales comme variables dépendantes, et les variables démographiques.

3.4. L'opérationnalisation (mise en œuvre)

Dans cette section, nous décrivons l'opérationnalisation des concepts en éléments concrets. Tout d'abord, nous décrivons les scénarios que nous avons développés pour les études, puis pour le concept d'agentivité et les variables, pour lesquelles nous décrivons les éléments et les questions correspondantes. Nous précisons également quelle échelle nous utilisons, et comment le concept d'agentivité est calculé. Nous décrivons dans les grandes lignes l'opérationnalisation de la vérification de l'attention, et indiquons quelles sont les variables démographiques que nous avons incluses dans l'étude.

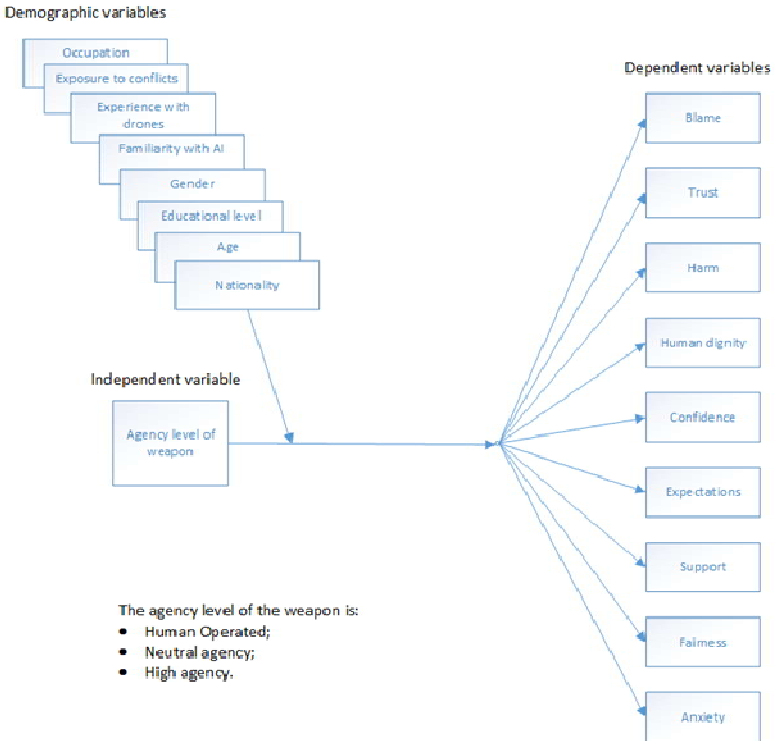


Figure 4 – Conception de recherche

3.4.1. Scénarios

Les scénarios sont utilisés dans le domaine de la science cognitive comme moyen d'étudier le jugement moral dans les expériences contrôlées randomisées (Cushman & Young, 2011; Kominski, Phillips, Gerstenberg, Lagnado & Knobe, 2015). Nous les avons utilisés pour étudier la perception de l'agentivité des armes autonomes, et avons créé un scénario qui décrit une opération militaire dans laquelle un convoi livre des approvisionnements dans une zone de conflit. Un véhicule

s'approche du convoi à grande vitesse; c'est une situation qui est susceptible de se présenter au cours de ce type d'opérations, et les personnels militaires doivent pouvoir estimer le niveau de menace pour décider ou non d'une attaque. Nous avons choisi de nous intéresser aux drones, car cette technologie est couramment utilisée par des opérateurs humains, et les drones sont déjà conçus pour avoir un degré d'autonomie par plusieurs compagnies, comme BAE Systems,[7] Dassault Aviation[8] et Boeing.[9] Bien que ces drones autonomes n'aient pas encore été déployés dans des opérations militaires, nous pensons que cela est susceptible d'arriver dans les cinq prochaines années.

Nous avons créé un scénario par défaut pour toutes les études que nous pouvions élargir, fondé sur les caractéristiques agentivité que nous voulions explorer. Voici un extrait du scénario par défaut que nous avons utilisé dans l'étude finale:

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais cause également des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

Ce scénario représente la condition d'agentivité neutre, dans laquelle nous ne donnons pas de caractéristiques

[7] https://en.wikipedia.org/wiki/BAE_Systems_Taranis
[8] https://en.wikipedia.org/wiki/Dassault_nEUROn
[9] https://en.wikipedia.org/wiki/Boeing_Phantom_Ray

d'agentivité pour l'arme. Dans les scénarios où l'homme est opérateur, nous remplaçons le terme drone autonome par drone actionné par l'homme, et ne donnons aucune information supplémentaire sur les caractéristiques agentivité (on notera que les mots sont surlignés en bleu pour faire ressortir les distinctions dans cette étude; les personnes interrogées dans l'enquête ont reçu des scénarios imprimés en noir). Dans la condition de haut degré d'agentivité, nous avons ajouté l'extrait suivant pour définir les caractéristiques agentivité: *«Le drone autonome choisit en toute indépendance dans une série d'options, pèse le pour et le contre, et décide d'attaquer le véhicule qui s'approche …»* Dans l'étude finale, nous avons précisé qu'il y a dommages collatéraux après l'attaque, mais dans les études pilotes nous avons aussi testé des scénarios sans dommages collatéraux, déclarant : *«qui cause la mort des quatre passagers, mais ne cause pas de dommages collatéraux.»*

Nous avons délibérément maintenu les modifications apportées aux scénarios au minimum de manière à pouvoir attribuer différents résultats à ces modifications et mesurer leur effet. Les résultats montrent que ces petits changements textuels conduisent à des différences dans la valeur moyenne des variables dépendantes et indépendantes ; il semble donc que le traitement consistant à utiliser les scénarios comme moyen d'effectuer une expérience contrôlée randomisée fonctionne bien.

Tableau 5 – Nombre de caractéristiques de l'agentivité rencontrées dans les publications

Caractéristique d'agentivité	**Nombre d'occurrences**	**Elément**
Prévoyance, pensée, réflexion, convictions, convictions	5	Pensée
Intentionnalité, intentions, désirs, désirs	4	Détermination
Planification, capacité à planifier, intentions, intentions	4	Buts atteints
Moralité, comprendre et appliquer des règles morales, normes	3	Règles morales
Contrôle de soi, liberté de choix	2	Agir librement
Autorégulation	1	
Introspection	1	
Autonomie	1	
Responsabilité	1	
Mémoire	1	
Emotion	1	
Reconnaissance	1	
Communication	1	
Capacité d'agir	1	

3.4.2. Concept d'agentivité

A partir de la revue des publications existantes sur les caractéristiques de l'agent dans la section 2.3., nous avons opérationnalisé le concept d'agentivité. Nous avons dressé une liste dans laquelle nous avons groupé les caractéristiques que nous jugions être semblables, et avons calculé le nombre de fois qu'une caractéristique était évoquée dans les publications (Ta-

bleau 5). Nous avons sélectionné les caractéristiques qui étaient évoquées plus d'une fois, et avons formulé une question pour les mesurer. Cela a eu pour résultat un concept d'agentivité composé initialement des cinq éléments qui étaient évoqués plus d'une fois dans les publications. Ces éléments seront mesurés sur une échelle d'auto-évaluation allant de 0 (ne sont pas du tout d'accord) à 100 (sont tout à fait d'accord), et la moyenne du score de l'agentivité sera calculée par rapport aux scores des éléments.

Nous avons utilisé les questions suivantes pour mesurer le concept d'agentivité:

1. CAPACITE DE REFLEXION (**PENSEE)**
   Le drone réfléchit de manière indépendante sur ce qu'il faut faire du véhicule, réfléchissant sur une série d'options pour défendre le convoi.

2. CAPACITE DE GENERER SES PROPRES OBJECTIFS (**DETERMINATION**)
   Le drone décide de manière indépendante si son objectif doit être d'éliminer le véhicule pour défendre le convoi.

3. CAPACITE D'APPLIQUER DES REGLES MORALES (**REGLES MORALES**)
   Le drone consulte de manière indépendante les règles et les normes morales fixées par le chef militaire, et décide s'il est moralement approprié d'éliminer le véhicule pour défendre le convoi.

4. CAPACITE D'AGIR DE SA PROPRE VOLONTE **(AGIR LIBREMENT**)

Le drone se trouve confronté à diverses possibilités et décide de manière indépendante s'il lui faut éliminer le véhicule pour défendre le convoi.

5. CAPACITE DE PLANIFIER POUR ATTEINDRE SON OBJECTIF (**ATTEINDRE SON OBJECTIF**)
   L'objectif du drone est de défendre le convoi; il décide donc de manière indépendante s'il lui faut initier un plan dans lequel le véhicule est une cible, dans lequel il faut choisir le type d'arme et lancer une attaque.

Après l'étude pilote 1, l'élément 3 (règles morales) a été abandonné, laissant pour le concept d'agentivité les quatre éléments qui ont été utilisés dans l'étude pilote 2 et l'étude finale:
1. Réflexion
2. Détermination de l'objectif
3. Liberté d'action
4. Réalisation de l'objectif

3.4.3. Variables dépendantes

Nous avons inclus 9 à 12 questions dans les études pilotes et l'étude finale pour explorer d'autres implications de la perception de l'agentivité des armes autonomes. Les variables dépendantes proviennent des résultats de l'étude des publications, du questionnaire sur les valeurs et des entretiens. Basée sur les résultats des études pilotes, l'étude finale a intégré les variables dépendantes; *imputation (blâme), confiance, tort causé, dignité humaine, assurance, attentes, approbation, équité et inquiétude.* Chacune de ces variables a été mesurée sur une échelle d'auto-évaluation allant de 0 (pas du tout d'accord) à 100 (entièrement d'accord).

1. IMPUTATION (BLAME): l'action est imputable au drone.

2. CONFIANCE: on peut faire confiance au drone pour prendre les mesures appropriées dans l'avenir.
3. TORT CAUSE: les actions du drone ont causé du tort.
4. DIGNITE HUMAINE: les actions du drone respectent la dignité humaine.
5. ASSURANCE: j'ai l'assurance que je drone prendra les mesures appropriées dans l'avenir.
6. ATTENTES: les actions du drone correspondent à mes attentes.
7. SOUTIEN / APPROBATION: j'approuve l'utilisation de ce type de drone par les militaires.
8. EQUITE: les actions du drone sont équitables.
9. INQUIETUDE: les actions du drone m'inquiètent.

Les concepts suivants ont été utilisés dans les études pilotes 1 et 2, mais abandonnés dans l'étude finale:

10. FAUTE DU CHEF: le chef est responsable de l'action.
11. CONFIANCE DONNEE AU CHEF: J'ai confiance en le chef pour qu'il prenne les mesures appropriées dans l'avenir.
12. TORTS CAUSES PAR LE CHEF: les actions du chef ont causé du tort.

3.4.4. Test d'attention

Nous avons ajouté une question avec un test d'attention au milieu des questions sur les variables dépendantes, dans laquelle nous avons demandé de choisir la valeur 40 pour cette question. Le test d'attention est souvent ajouté à un questionnaire dans les études de psychologie cognitive, par exemple par Logg (2017), et les personnes qui ne donnent pas la réponse correcte sont exclues de l'échantillon, considérant qu'elles ne sont pas non plus suffisamment attentives pour les autres questions.

3.4.5. Variables démographiques

Nous avons également ajouté des questions pour recueillir des informations d'ordre «démographique» sur les personnes interrogées, et avons ajouté des variables sur: *l'âge, le genre, le niveau d'études, la profession et la nationalité.* Après ces questions générales de nature «démographique», nous avons demandé aux personnes interrogées si elles avaient une quelconque *expérience de l'intelligence artificielle*, si elles *travaillaient avec des drones* et si elles s'étaient déjà *trouvées dans une zone de conflit.*

3.5. Approche analytique

Cette section décrit l'approche que nous avons adoptée pour analyser les données. Pour chacune des études nous avons procédé en suivant les phases décrites ci-dessous dans l'analyse des données.

3.5.1. Prétraitement des données

Dans l'outil Qualtrics après l'achèvement de l'étude, une distinction est faite entre *réponses en cours* et *réponses complètes*. Avant de commencer l'analyse des données, le pourcentage de finalisation des réponses en cours doit être vérifié et on doit décider si elles doivent être regroupées avec les réponses complètes. Il s'avère que lorsqu'Amazon MTurk est utilisé pour diffuser l'enquête, bon nombre de personnes interrogées oublient de cliquer sur le dernier bouton, ce qui fait que 99% de leurs réponses sont complètes, mais ces réponses sont valables et utilisables. Si la personne interrogée a répondu à toutes les questions, nous les avons ajoutées à la liste des réponses complètes. Si des personnes interrogées n'ont pas répondu à toutes les questions, par exemple aux questions d'ordre démographique, nous les avons éliminées de l'échantillon. Dans la phase suivante de prétraitement, nous avons éliminé toutes les personnes qui ont échoué au test d'attention auquel nous procé-

dons à mi-chemin dans le questionnaire. Nous avons demandé si les personnes ont répondu 40 à cette question, et nous avons éliminé toutes les réponses inférieures à 38 et supérieures à 42; nous supposons que cela révélait une erreur de précision pour placer les curseurs.

3.5.2. Analyse de fiabilité

Après avoir téléchargé le dossier csv dans SPSS (logiciel d'analyse statistique), nous avons procédé à un test de fiabilité qui est une mesure pour évaluer la cohérence interne d'un ensemble d'éléments de test ou d'échelle. Le test de fiabilité vérifie si le mesurage donné est cohérent pour un concept. Le système qui mesure la cohérence interne est Alpha de Cronbach, qui doit être supérieur à 0.7 pour démontrer une cohérence interne du concept. Pour chacune des études l'Alpha de Cronbach pour les éléments du concept d'agentivité est donné, car ces éléments sont utilisés pour mesurer le concept d'agentivité, et doivent avoir une cohérence interne. On indique également si l'Alpha de Cronbach est amélioré si l'un des éléments est enlevé. La fiabilité des variables dépendantes est vérifiée, mais pas communiquée, car ces éléments ne sont pas conçus pour être utilisés comme une seule échelle, et il s'est avéré que leur cohérence interne est plus basse que le seuil ($\alpha<0.7$), comme on pouvait s'y attendre.

3.5.3. Analyse des principales composantes (ACP)

Dans la phase suivante de l'analyse des données, une analyse de la composante principale (ACP) a été menée. Cette phase est une technique de réduction de variable dont le but est de réduire un plus grand ensemble de variables en un plus petit ensemble, appelé «composantes principales», qui expliquent la plupart des variations dans les variables d'origine. Nous avons procédé à l'ACP sur les éléments d'agentivité, pour voir s'ils

pouvaient être agrégés en un seul concept d'agentivité, et indiquons la variation et le nombre des composantes comme résultats. Nous avons aussi vérifié si le concept était influencé par les variables démographiques, mais ne communiqueront pas ces résultats, comme l'ACP indique que toutes les variables démographiques sont des composantes distinctes qui n'influencent pas le concept d'agentivité.

3.5.4. Analyse de corrélation

La Corrélation Pearson bivariée (à deux variables) est un test pour mesurer la force et la direction des relations linéaires entre deux variables continues. Elle produit un coefficient de corrélation *r* qui peut varier entre -1 et +1. Une valeur négative indique une relation négative entre les concepts, par exemple plus la perception de l'agentivité est grande, moins l'est l'approbation (soutien). Une valeur positive indique une relation positive, par exemple plus la perception de l'agentivité est grande, plus l'élément imputation (ou blâme) l'est aussi. Un coefficient de relation égal à zéro indique qu'il n'y a pas de relation entre les concepts. Nous avons procédé à cette analyse pour vérifier la corrélation entre le concept d'agentivité et les variables dépendantes, et indiquons les résultats.

3.5.5. Vérification concernant la manipulation sur l'agentivité

Le test de manipulation vérifie si le concept d'agentivité varie entre les scénarios, et par là nous avons vérifié si les hypothèses devaient être retenues ou rejetées. La vérification se fait de deux manières. D'abord on crée un graphique, le concept d'agentivité apparaît sur l'axe des y, et les scénarios sur l'axe des x, pour l'analyse visuelle. On s'attend à ce que les niveaux d'agentivité soient différents selon les conditions considérées, et cette différence devrait être de l'ordre de $p < 0.5$. Ceci est indiqué par les barres en forme de I en tête de colonne et

montre l'erreur moyenne multipliée par 1, et indique si les concepts se chevauchent ou non. Si les barres se chevauchent, alors les deux groupes ne sont pas significativement différents.

Le même résultat peut être obtenu par des tests T (test de Student) d'échantillons indépendants, pour vérifier s'il existe des différences importantes dans les moyennes de deux groupes distincts, pour lesquels les questions sont attribuées au hasard, si bien que l'effet observé est le résultat du traitement (dans ce cas lire un scénario spécifique), et ne peut pas être attribué à un autre effet. Nous présenterons les graphiques et communiquerons les résultats du test T concernant des échantillons indépendants, avec les différents niveaux d'agentivité des armes autonomes.

3.5.6. Analyse des variables dépendantes

L'analyse des variables dépendantes est de nature exploratoire, et les résultats sont présentés sous forme de graphique. Chaque graphique indique la moyenne de la variable dépendante sur l'axe des y, et les différents scénarios sur l'axe des x pour faire ressortir la différence entre les scénarios. Bien que la description des variables dépendantes soit descriptive, nous indiquons les résultats en comparant les différentes conditions significatives au niveau $p < .05$.

3.6. Préinscription

L'étude finale a été préenregistrée sur l'Open Science Framework[10] qui crée une version marquée de la date d'un projet qui ne peut pas être supprimé ou édité. Bien que ce ne soit pas encore exigé des publications scientifiques, les analystes vérifient souvent si le projet est préenregistré. Nous avons utilisé le modèle AsPredicted dans lequel on répond à 8 questions concer-

[10] https://osf.io/

nant l'étude, par exemple si les données ont déjà été rassemblées, quelles sont les hypothèses qui sont testées, et le type d'analyses que l'on a l'intention d'effectuer. Le pré enregistrement est placé «sous embargo» jusqu'au 24 juin 2021. Nous avons pré enregistré la solide hypothèse que nous avons sur la perception de l'agentivité, et pensons pouvoir faire part des constatations faites à la suite des résultats de l'étude pilote 2. Pour ce qui est des dépendantes variables, nous ne sommes pas sûrs de savoir quels mécanismes sont en jeu, ni si nous pourrons communiquer nos résultats ; par conséquent nous avons enregistré une analyse exploratoire de ces résultats.

3.7. Echantillon

Toutes les enquêtes ont été réalisées sous Qualtrics, et les études pilotes ont été diffusées par la plateforme de crowdsourcing (externalisation ouverte) Amazon Mechanical Turk. Nous avons testé l'enquête avec un petit échantillon de 20 personnes interrogées, et examiné le temps moyen pour compléter les réponses, pour vérifier si nous avions estimé les délais de complétion correctement. Nous avons vérifié que les scénarios étaient diffusés de manière égale, si les personnes vérifiaient la question test correctement, et s'il y avait des remarques au sujet de l'enquête. Comme tout avait été bien vérifié, nous avons décidé de porter l'échantillon représentatif à 50 personnes par scénario, ce qui donnait 500 participants pour l'étude pilote 1 et 700 pour l'étude pilote 2. Le recueil des réponses prit environ 4 heures, et chaque personne interrogée perçut 0.40 dollar US.

L'étude finale a été diffusée via la méthode «boule de neige» par email avec un lien anonyme vers environ 40 collègues des armées qui à leur tour propagèrent l'enquête. Cette méthode a été utilisée car nous n'avions pas l'autorisation du MIT ni du Ministère de la Défense néerlandais de recueillir des

informations personnelles comme des emails ou adresses internet. En utilisant la méthode «boule de neige», nous n'avions aucune garantie quant au nombre des personnes interrogées. Nous avons également publié une annonce sur le réseau interne de l'Armée de Terre. L'étude finale s'est déroulé sur 16 jours, du 12 juin au 27 juin 2017, et a produit 327 réponses, dont 239 étaient des réponses valides et complètes, et utilisables après le prétraitement des données.

3.8. Questions méthodologiques

Dans cette section, plusieurs questions et problèmes concernant la méthodologie seront abordés. Ces problèmes pourraient avoir des implications pour la validité interne et externe de la présente étude et lorsque c'est possible, quelques contremesures sont données pour traiter ces problèmes.

3.8.1. Entretiens pour le codage

Pour mieux connaître les valeurs des experts nous avons procédé à des entretiens semi-structurés; c'est une méthode de recherche qualitative. Nous avons utilisé la méthode de Codage des Valeurs (Value Coding method) pour interpréter les données et extraire les valeurs des entretiens après transcription. L'un des problèmes qui se posent lors du codage de données quantitatives est que c'est une activité heuristique, et aucune formule pour le codage n'existe. Il faut aussi trouver des liens et donner un sens aux données (Saldana, 2015). Ces caractéristiques impliquent que la méthode de codage des valeurs est susceptible d'être biaisée, car le chercheur utilise son propre cadre et ses notions personnelles pour interpréter les données. Pour réduire cette distorsion en appliquant la méthode d'encodage des valeurs un second chercheur codera les entretiens indépendamment du premier. Des instructions sur la mé-

thode seront données, mais peu d'informations sur les valeurs et les résultats.

3.8.2. Expériences contrôlées randomisées (études expérimentales avec randomisation)

L'une des conditions pour procéder à des expériences contrôlées randomisées est, comme leur nom l'indique, la diffusion des scénarios auprès des personnes interrogées. La randomisation limite la confusion, renforce la validité interne de la recherche, et constitue un prérequis indispensable pour ce type d'étude. L'un des problèmes survient lorsque la randomisation échoue et que la répartition des répondants est faussée du fait des scénarios. L'une des causes de diffusion faussée peut être que le logiciel ne répartit pas de façon égale les répondants (personnes interrogées) sur les scénarios. Les chercheurs ne peuvent que vérifier si les programmes du logiciel sont corrects. Une autre cause de diffusion faussée est que des personnes voient un certain scénario qui ne leur plaît pas ou auquel ils ne s'attendaient pas, et quittent l'enquête avant de finir. Par exemple, des personnes s'attendent à répondre à une enquête sur les armes autonomes et peuvent se voir attribuer un scénario sur une arme actionnée par l'homme, et alors décident d'arrêter. Le fait de quitter l'enquête prématurément est appelée «attrition sélective», et cela peut résulter en une distorsion dans le travail de recherche, car les personnes qui abandonnent sont fondamentalement différentes de celles qui y participent. Dans le cadre actuel de la recherche, il n'est pas possible de prendre des mesures pour y remédier, et ceci peut constituer une menace pour la validité interne de l'étude.

3.8.3. Amazon Mechanical Turk

L'avantage qu'il y a à utiliser Amazon MTurk comme méthode de diffusion d'une étude est qu'un grand nombre de répon-

dants participeront à l'enquête et répondront aux questions avec sérieux pour conserver leur statut professionnel. La méthode a également des inconvénients car les collaborateurs sur MTurk sont surtout basés aux USA, du fait que pour être employé vous devez vous plier aux règles américaines concernant les impôts même si ne résidez pas aux Etats-Unis. Ceci pourrait conduire à une distorsion dans les résultats, impliquant que les résultats ne sont pas généralisables en dehors des USA. Bonnefon et al (2016) signalent un deuxième problème dans l'utilisation de MTurk: les participants peuvent ne pas être représentatifs de la population US, par exemple parce qu'un grand nombre des collaborateurs sont des étudiants, et un troisième problème est que les collaborateurs connaissent déjà bien ce dont il s'agit. Le premier problème peut être résolu en choisissant un large éventail de collaborateurs, si bien que l'échantillon comptera également des personnes ne résidant pas aux USA. Le deuxième problème est inhérent au choix de MTurk et il n'existe pas de mesures à prendre pour y remédier. Le troisième problème n'est pas vraiment susceptible de survenir, car les scénarios sont uniques et n'ont fait l'objet d'étude sur MTurk auparavant.

3.8.4. Enquête finale, diffusion «boule de neige»

L'étude finale a été diffusée aux collègues militaires qui ont transmis leur email à travers leur réseau. Utiliser la méthode de diffusion «boule de neige» («snowball») présente quelques inconvénients. Le premier est que les participants pouvaient participer à l'enquête plusieurs fois, et du fait que nous ne sommes pas autorisés à stocker des informations d'identité personnelle, on ne peut pas vérifier si c'est le cas. Ceci pouvait conduire à une distorsion dans les résultats si les répondants voyaient des scénarios différents, et influencer les réponses. Un deuxième inconvénient est que des collègues répondent au test les pre-

miers, puis font part de leur expérience aux autres participants avant que ceux-ci répondent. Ceci influencerait également leurs réponses. En troisième lieu, les chercheurs ne peuvent pas contrôler qui répondra à l'enquête; par conséquent, un risque potentiel est que l'enquête soit transmise à quelqu'un qui ne travaille pas au Ministère de la Défense, et que l'échantillon soit «contaminé». Malheureusement, comme on ne peut pas enregistrer des informations d'identité personnelle, il n'y a pas beaucoup de mesures possibles pour remédier à ces trois problèmes.

4. Résultats

> *On peut espérer que les gens se rendent compte que les ordinateurs feront uniquement ce qu'ils ont été programmés à faire, et RIEN d'autre.*
>
> Pilote interrogé, étude 1

Cette section donne les résultats de l'enquête sur les valeurs, qui consistait en un questionnaire et quelques entretiens d'experts, ainsi que les expériences contrôlées randomisées, incluant deux études «pilotes» et l'étude finale. Les valeurs obtenues à partir du questionnaire en ligne et des entretiens sont fusionnées en une seule présentation. Les résultats des études «pilotes» et de l'étude finale sont présentés en utilisant la structure de phases d'analyse de données (section 3.5.) et les principaux résultats sont donnés dans la conclusion de chaque étude.

4.1. Enquête sur les valeurs

Pour connaître quelles sont les valeurs que les personnes associent aux armes autonomes, une enquête et six entretiens d'experts ont été menés. Un aperçu des résultats est donné dans ce chapitre.

4.1.1. Enquête «en ligne»

L'enquête sur les valeurs consistait en un questionnaire demandant aux répondants quelles valeurs de bioéthique et de celles de Cheng et Fleischmann s'appliquent. Dans la question finale, on demandait aux participants de donner au moins une valeur supplémentaire qui n'était pas évoquée dans les questions de départ. Au total 69 répondants ont répondu à l'enquête, et après avoir éliminé les réponses incomplètes, il restait 57 ré-

ponses complètes. Les résultats sont détaillés dans les graphiques ci-dessous.

Le court questionnaire permet de savoir quelles valeurs les personnes associent aux armes autonomes. Les résultats concernant la question bioéthique montrent que la *non-malfaisance* (non-nuisibilité) est la plus importante, ensuite vient la *«bienfaisance»* (capacité à faire le bien), en troisième la *justice*, et enfin l'*autonomie*. Les cinq valeurs de tête sur la base de la liste de Cheng et Fleischmann (2010) sont: *sécurité, intelligence, responsabilité, justice* et *ordre social*. Les questions ouvertes révèlent que les personnes associent les armes autonomes surtout à l'*imputabilité* et à *l'efficacité*.

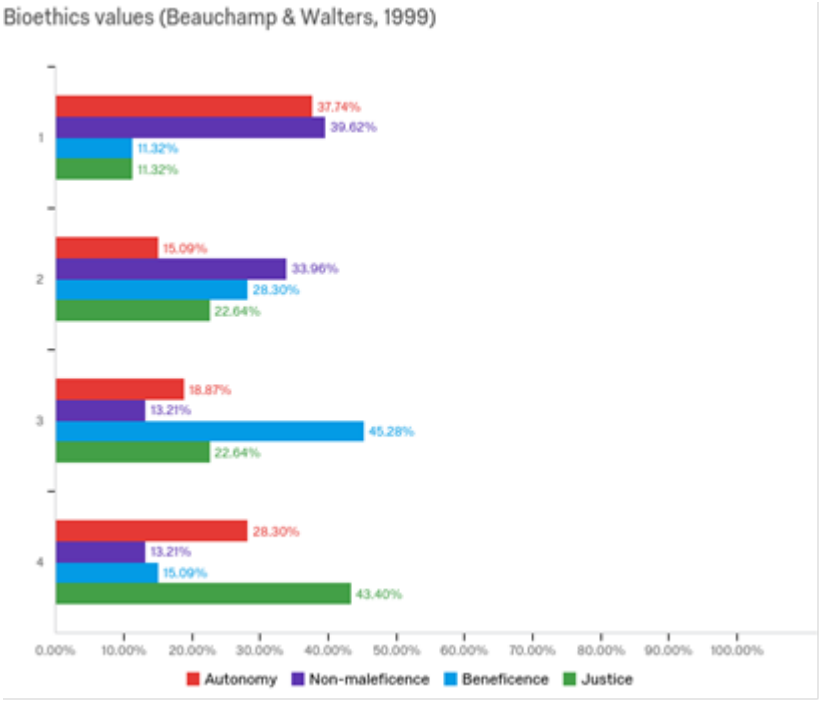


Figure 5 – Résultats question 1, enquête «en ligne» sur les valeurs

Figure 6 – Résultats question 3, enquête «en ligne» sur les valeurs

4.1.2. Entretiens

Six entretiens ont été conduits avec des experts dans les domaines suivants:

INTELLIGENCE ARTIFICIELLE:

- Prof. Toby Walsh (université de Nouvelle Galles du Sud & Data61).

ARMES AUTONOMES:

- Dr. Peter Asaro (the New School & Campaign Ban Killer Robots, campagne pour l'élimination des robots tueurs);
- Mme. Miriam Struyk (directeur des programmes, Security & Disarmament PAX);
- Dr. Michael Horowitz (Université de Pennsylvanie).

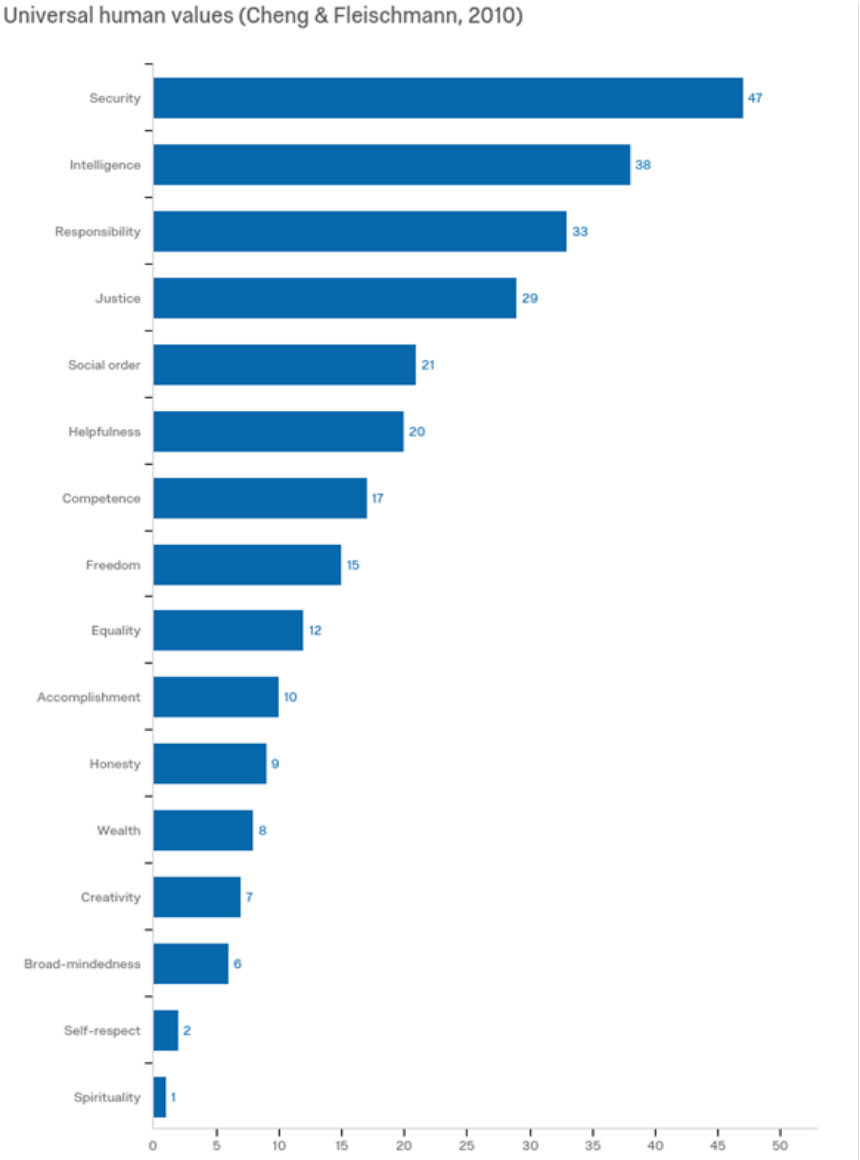


Figure 7 – Résultats question 2, enquête «en ligne» sur les valeurs

OPERATIONS MILITAIRES:

- Lieutenant-Colonel Kremers (Etat-major Armée de Terre néerlandaise);
- Capitaine de corvette van den Sande (Etat-major Ministère de la Défense).

Après les discussions sur les valeurs, les entretiens ont fourni de précieuses informations de fond sur l'histoire et les positions des différents groupes dans le cadre du débat actuel sur les armes autonomes, mais ont aussi montré qu'il n'existe pour le moment aucun consensus sur une définition. La transcription complète de ces entretiens figure dans l'Appendice F, Transcriptions d'entretiens.

Les entretiens ont été encodés au moyen de la méthode «encodage des valeurs», en se basant sur la comparaison faite entre l'encodage du chercheur et celui du réviseur (Appendice H, processus d'encodage des résultats), la liste suivante de 22 valeurs uniques a été dressée:

1. Intervenir n'importe quand
2. Les rappeler
3. Minimiser les risques
4. Avoir le contrôle
5. Bonne volonté
6. Fixer des limites
7. Droit d'être tué par une personne
8. Cadre éthique
9. Contrôle humain significatif
10. Prévisibilité
11. Réflexivité
12. Distance
13. Invincibilité
14. Déterminisme
15. Inévitabilité
16. Souffrance inutile
17. Blessure superflue
18. Dignité humaine
19. Imputabilité
20. Responsabilité
21. Défense
22. Tort causé

4.1.3. Enquête sur les valeurs: conclusion

Pour conclure sur l'enquête sur les valeurs, nous comparerons les valeurs ressortant de l'enquête en ligne et les entretiens qui donnent lieu à une liste de valeurs (Tableau 6). Six valeurs dans cette liste sont mentionnées à la fois dans l'enquête sur les valeurs et dans les entretiens. A partir des questions que les experts ont mises en exergue au cours des entretiens, nous avons sélectionné les valeurs qui étaient le plus souvent évoquées pour les tester dans l'étude pilote 1. Ces valeurs sont *l'équité, le tort causé, la dignité humaine* et *la responsabilité.*

Enquête «en ligne»	**Entretiens**
Justice	
Autonomie	
Sécurité	
Intelligence	
Ordre social	
Efficacité	
Equité	
Racisme	
Contrôle	En situation de contrôle
Capacité à faire le bien	Bonne volonté
Imputabilité	Imputabilité
Responsabilité	Responsabilité
Sûreté	Sûreté
Non-malfaisance	Tort causé
	Capacité à intervenir à tout moment
	Les rappeler
	Minimiser les risques
	Fixer des limites
	Droit d'être tué par une personne
	Cadre éthique

	Contrôle humain significatif
	Prévisibilité
	Réflexivité
	Distance
	Invincibilité
	Déterminisme
	Inévitabilité
	Souffrance inutile
	Blessure superflue
	Dignité humaine

Tableau 6 – Valeurs ressortant de l'enquête en ligne et des entretiens

4.2. Etude Pilote 1

Dans la première étude pilote, nous avons essayé de vérifier si nous pouvions manipuler la perception de l'agentivité des drones actionnés par l'homme et des armes autonomes. Nous nous sommes intéressés à la manière dont les gens percevaient la technologie actuelle comparée à la technologie du futur. Nous avons créé plusieurs conditions dans lesquelles nous avons varié le niveau d'agentivité en donnant des informations supplémentaires sur la manière dont le drone «réfléchissait» sur la possible attaque d'un véhicule, ou en laissant délibérément de côté ces informations. Nous avons testé les conditions «bon résultat» et «mauvais résultat», car nous pensions que cela aurait un impact sur la manière dont les variables dépendantes, par exemple l'approbation / soutien, seraient considérées. Dans cette étude, nous avons posé de manière spécifique les questions ayant trait à l'agentivité des drones et non à l'opérateur humain qui les contrôle.

Nous avons testé 10 scénarios, avec le but d'obtenir 50 répondants par condition, et donc un total de 500 répondants. Le nombre de répondants par scénario s'élevait de 49 à 54. Après le prétraitement des données, et après avoir éliminé les

répondants qui avaient échoué au test d'attention, il nous restait 510 réponses complètes.

	Condition d'agentivité				
	Drone autonome			**Drone avec opérateur humain**	
	Agentivité zéro	Agentivité basse	Agentivité élevée	Agentivité basse	Agentivité élevée
Pas de dommages collatéraux (bon résultat)	1	2	3	4	5
Dommages collatéraux (mauvais résultat)	6	7	8	9	10

Tableau 7 – Scénarios de l'étude pilote 1

4.2.1. Analyse de fiabilité

Pour les cinq éléments d'agentivité $\alpha = 0.9$, ce qui est bien au-dessus du seuil de 0.7, et ne peut être amélioré que si l'élément SQ3 Règles morales est éliminé comme cela apparaît dans le Tableau 8. La suppression d'autres éléments diminuera la valeur α et réduira la cohérence interne, comme on le voit dans la dernière colonne (si l'élément Alpha de Crohnbach est supprimé) dans le Tableau 8.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.900	.895	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
SQ1_Thought	167.77	13710.694	.841	.741	.857
SQ2_Goal setting	163.85	13440.002	.857	.775	.853
SQ3_Moral rules	183.39	17599.970	.448	.215	.933
SQ4_Act freely	163.22	13885.447	.824	.695	.861
SQ5_Achieve goals	157.21	13891.495	.794	.681	.868

Tableau 8 – Résultats analyse de fiabilité éléments d'agentivité – étude pilote 1

4.2.2. Analyse composante principale (ACP)

L'ACP montre que SQ1 Réflexion, SQ2 Détermination des objectifs, SQ4 Liberté d'action, et SQ5 Réalisation des objectifs peuvent être considérés comme un ensemble qui s'élève à 71.75% des variables parmi les variables d'origine (Tableau 9). On peut aussi le constater dans le diagramme d'éboulis dans lequel apparaissent les valeurs propres des composantes (Figure 8). A partir de la matrice des composantes, nous pouvons observer que SQ3 Valeurs morales est une composante à part en comparaison des quatre autres éléments d'agentivité.

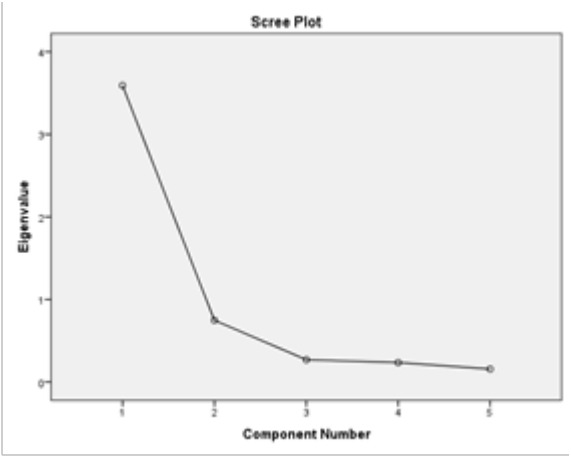


Figure 8 – Diagramme d'éboulis ACP, Etude pilote 1

Component Matrix[a]

	Component		
	1	2	3
SQ2_Goal setting	.922		
SQ1_Thought	.911		
SQ4_Act freely	.899		
SQ5_Achieve goals	.882		
SQ3_Moral rules	.571	.819	

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Tableau 9 – Eléments d'agentivité, Etude pilote 1

4.2.3. Analyse corrélation

La corrélation entre le concept (ensemble) agentivité et les variables dépendantes est petite, et significative seulement pour DVQ1 *Faute / blâme*, DVQ2 *Faute du chef*, DVQ4 *Confiance en le chef*, DVQ6 *Tort causé par le chef* et DVQ8, *Approbation / Soutien*.

Quelques effets inattendus sont observables du côté des relations. Par exemple, on peut s'attendre à ce que la perception de l'agentivité fait que les personnes attribuent plus de *faute / blâme* au drone (r = .183) mais ces résultats indiquent que les personnes attribuent moins de *faute / blâme* au chef (r = -.239). Le même effet peut être observé avec la variable tort causé. L'analyse de corrélation n'est pas assez détaillée pour se focaliser sur ces résultats, et donc cela sera fait dans l'analyse des variables dépendantes.

4.2.4. Vérification manipulation sur l'agentivité

Le graphique de la Figure 9 montre que notre manipulation sur l'agentivité fonctionne, car dans la condition de niveau bas d'agentivité pour les armes autonomes, les personnes perçoivent moins d'agentivité que dans la condition d'agentivité élevée, comme le montre la moyenne du concept d'agentivité sur l'axe des y. Dans la condition d'agentivité neutre la perception est un peu plus élevée que dans la condition de basse agentivité, mais plus basse que pour la condition d'agentivité élevée. Le test-T (test de Student) des échantillons indépendants montre que l'agentivité basse et l'agentivité élevée (armes autonomes) sont des ensembles distincts (p = .000), mais ce n'est pas le cas pour l'agentivité neutre et l'agentivité basse (armes autonomes) (p = .209), car la différence entre ces groupes n'est pas significative, ce qui peut être aussi observé dans le chevauchement des barres d'erreur pour ces conditions.

Le drone actionné par l'homme est perçu comme ayant une agentivité beaucoup plus basse que dans les conditions pour l'arme autonome. La manipulation de l'agentivité fonctionne même si la différence est moins évidente que dans le scénario des armes autonomes. Les barres d'erreur dans les scénarios à basse agentivité indiquent que les ensembles «drone

actionné par l'homme» ne peuvent être distincts, ce qui est confirmé par le test-T de l'échantillon indépendant (p = .125).

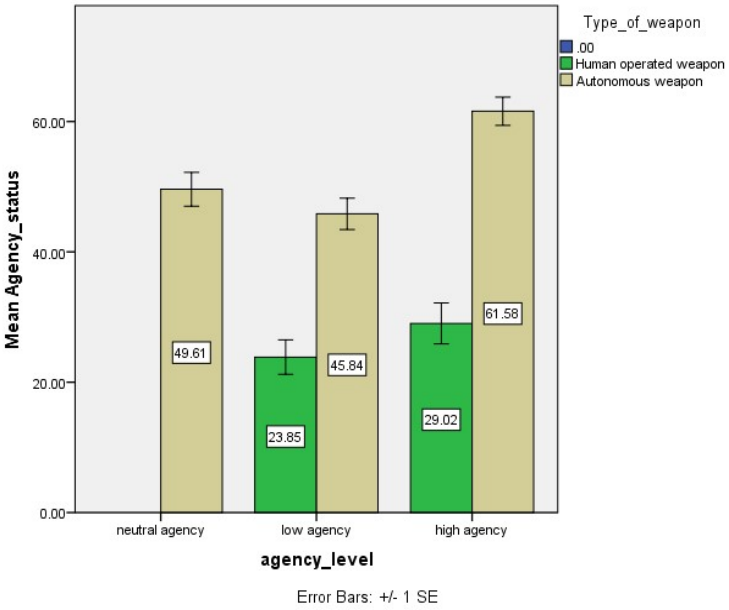


Figure 9 – Valeur moyenne concept agentivité par type d'arme

Le deuxième graphique (Figure 10) montre la moyenne de la perception de l'agentivité pour les niveaux d'agentivité, pour les «bons» et «mauvais» résultats. Le type de résultat de l'utilisation de l'arme, que ce soit des dommages collatéraux ou non, ne semble pas aboutir à beaucoup de différences dans la perception de l'agentivité, les moyennes variant de 2 à 3 points (sur une échelle de 100). Les barres d'erreur se chevauchent aussi sur tous les niveaux d'agentivité, ce qui est confirmé par les tests-T sur l'échantillon indépendant. Dans la condition niveau d'agentivité neutre, la différence entre les ensembles de bon résultat et de mauvais résultat est p = .816, dans les conditions de

basse agentivité elle est p = .641, et dans les conditions de haute agentivité de p = .575, toutes beaucoup plus élevées que le seuil de p < .05.

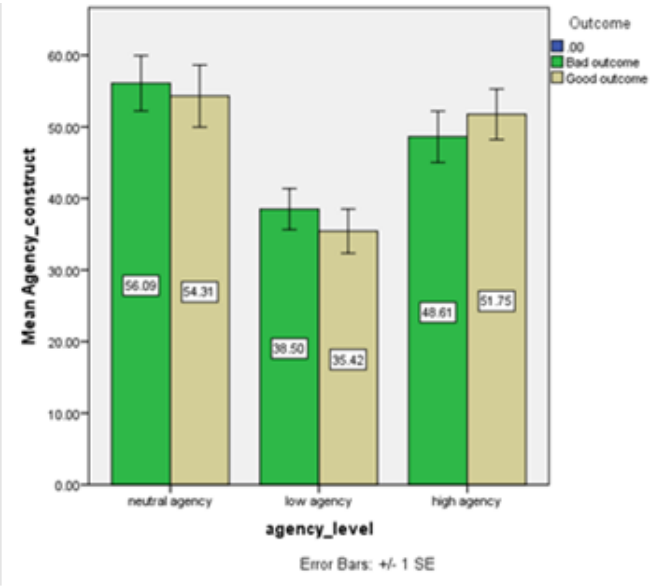


Figure 10 – Valeur moyenne du concept agentivité par condition (de niveau d'agentivité) par résultat

4.2.5. Analyse des variables dépendantes

L'analyse de corrélation a montré que seules les variables dépendantes *faute (blâme), faute du chef, confiance en le chef, tort causé par le chef* et *approbation / soutien* étaient significatives à un niveau minimum de p < .05. Dans l'étape suivante, nous avons regardé de plus près ces variables et du fait de leur non signifiance, nous n'avons pas analysé le reste des variables dépendantes. En

Correlations

		Agency_construct	DVQ1_Blame	DVQ2_CommanderBlame	DVQ3_Trust	DVQ4_CommanderTrust	DVQ5_Harm	DVQ6_CommanderHarm	DVQ7_Uneasy	DVQ8_Support	DVQ9_Fairness
Agency_construct	Pearson Correlation	1	.183**	-.239**	.070	.176**	-.011	-.227**	.005	.101*	.087
	Sig. (2-tailed)		.000	.000	.114	.000	.798	.000	.914	.023	.050
	N	510	510	510	510	510	510	510	510	510	510
DVQ1_Blame	Pearson Correlation	.183**	1	-.101*	.020	.053	.198**	-.170**	.129**	-.120**	-.078
	Sig. (2-tailed)	.000		.023	.660	.232	.000	.000	.003	.006	.077
	N	510	510	510	510	510	510	510	510	510	510
DVQ2_CommanderBlame	Pearson Correlation	-.239**	-.101*	1	-.264**	-.330**	.264**	.692**	-.001	-.333**	-.036
	Sig. (2-tailed)	.000	.023		.000	.000	.000	.000	.982	.000	.422
	N	510	510	510	510	510	510	510	510	510	510
DVQ3_Trust	Pearson Correlation	.070	.020	-.264**	1	.635**	-.294**	-.262**	.169**	.674**	-.106*
	Sig. (2-tailed)	.114	.660	.000		.000	.000	.000	.000	.000	.016
	N	510	510	510	510	510	510	510	510	510	510
DVQ4_CommanderTrust	Pearson Correlation	.176**	.053	-.330**	.635**	1	-.171**	-.322**	.138**	.584**	.021
	Sig. (2-tailed)	.000	.232	.000	.000		.000	.000	.002	.000	.630
	N	510	510	510	510	510	510	510	510	510	510
DVQ5_Harm	Pearson Correlation	-.011	.198**	.264**	-.294**	-.171**	1	.361**	-.044	-.336**	.089*
	Sig. (2-tailed)	.798	.000	.000	.000	.000		.000	.323	.000	.046
	N	510	510	510	510	510	510	510	510	510	510
DVQ6_CommanderHarm	Pearson Correlation	-.227**	-.170**	.692**	-.262**	-.322**	.361**	1	-.038	-.322**	.041
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.394	.000	.351
	N	510	510	510	510	510	510	510	510	510	510
DVQ7_Uneasy	Pearson Correlation	.005	.129**	-.001	.169**	.138**	-.044	-.038	1	.127**	-.689**
	Sig. (2-tailed)	.914	.003	.982	.000	.002	.323	.394		.004	.000
	N	510	510	510	510	510	510	510	510	510	510
DVQ8_Support	Pearson Correlation	.101*	-.120**	-.333**	.674**	.584**	-.336**	-.322**	.127**	1	-.004
	Sig. (2-tailed)	.023	.006	.000	.000	.000	.000	.000	.004		.934
	N	510	510	510	510	510	510	510	510	510	510
DVQ9_Fairness	Pearson Correlation	.087	-.078	-.036	-.106*	.021	.089*	.041	-.689**	-.004	1
	Sig. (2-tailed)	.050	.077	.422	.016	.630	.046	.351	.000	.934	
	N	510	510	510	510	510	510	510	510	510	510

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Tableau 10 Matrice de corrélation, concept d'agentivité et variables dépendantes

examinant les détails nous avons vérifié s'il y avait des différences entre les niveaux d'agentivité bas, neutre et élevé et entre les conditions de niveau pour les armes autonomes et les drones actionnés par l'homme.

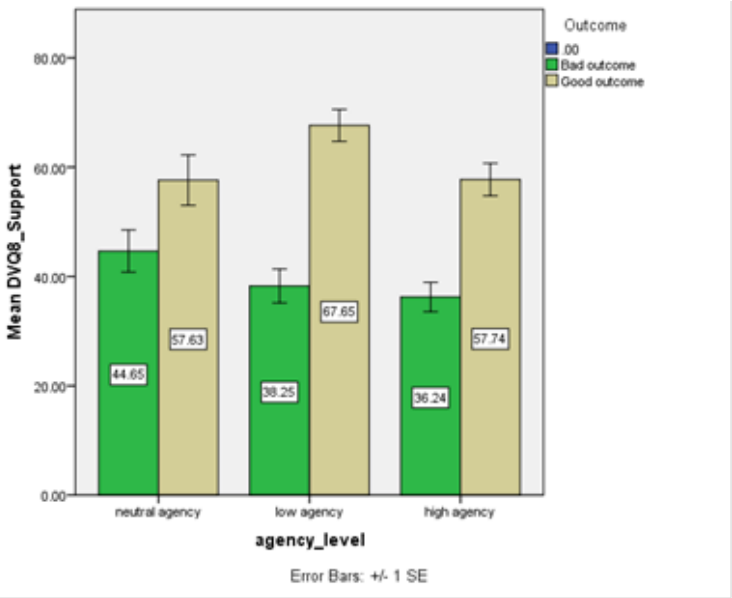


Figure 11 – Valeur moyenne de la variable «approbation» par condition (de niveau d'agentivité) par résultat

Variable approbation

L'observation la plus intéressante est la grande différence pour ce qui est de l'«approbation» (soutien) entre «bons» et «mauvais» résultats (Figure 11), qui est importante dans les trois conditions (de niveau d'agentivité) ($p < .05$). Cette différence substantielle dans la moyenne de l'«approbation» pour les trois conditions indique que notre manipulation au niveau des résultats fonctionne. Une autre observation intéressante est que

l'approbation diminue à mesure que les niveaux d'agentivité augmentent, mais les différences dans la moyenne sont petites (< 10 points sur une échelle de cent), et les barres en I en haut avec l'erreur standard en chevauchement indiquent que ces ensembles ne sont pas significativement différents, lorsqu'on les compare dans les scénarios à «bons» et «mauvais» résultats. Le second graphique montre que l'approbation est plus forte dans les conditions «opérateur humain» comparée aux conditions «arme autonome», mais là aussi les différences sont moindres et non significatives comme les barres d'erreur se chevauchent.

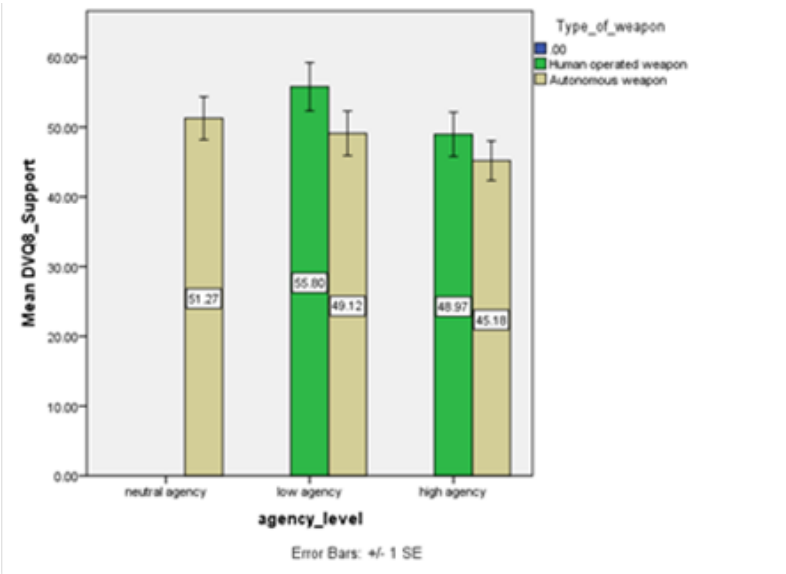


Figure 12 – Moyenne variable valeur «approbation» par condition (de niveau d'agentivité) par type d'arme

Valeur «Confiance»

Sur le graphique de la Figure 12 on peut voir que les personnes sont moins susceptibles d'avoir confiance en une arme auto-

nome pour prendre les mesures adéquates dans le futur après un mauvais résultat qu'en un opérateur humain qui prend des mesures avec le même mauvais résultat, et tous les résultats sont significatifs. Ce phénomène est magnifié dans les conditions de bas niveau d'agentivité où l'on n'observe pratiquement aucune différence entre la confiance entre un humain après un résultat bon ou mauvais, et une grande différence dans la confiance pour les armes autonomes (Figure 13). Cependant, l'analyse à une seule variable a montré qu'il n'y a pas d'interaction significative: type d'arme * résultat p = .120, ce qui signifie que l'on ne peut tirer aucune conclusion à ce point de l'étude, mais il est intéressant de continuer l'étude.

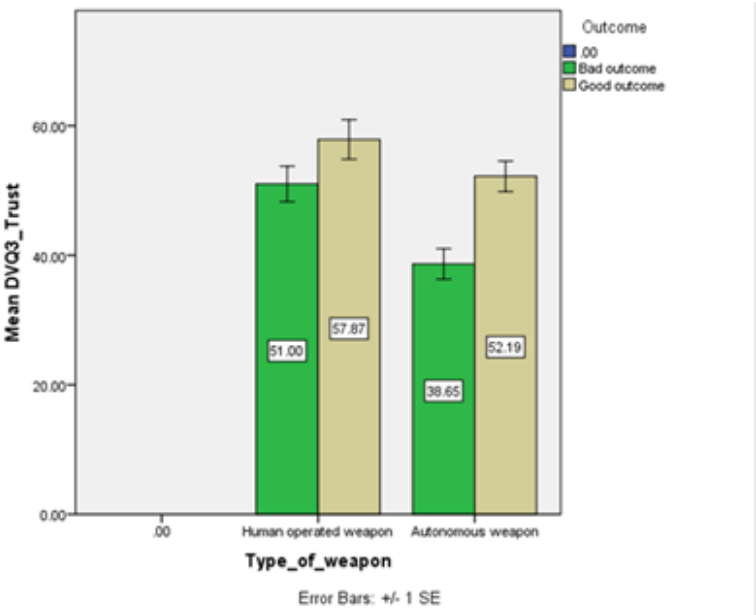


Figure 13 – Moyenne de variable «confiance» par type d'arme par résultat

Variables «faute» (blâme) et «tort» (causé)

Pour étudier la chaîne de responsabilité la moyenne des variables *faute, faute du chef, tort et tort causé par le chef* est présentée sur le même graphique (Figure 15). Tous les résultats des variables «faute» (Figure 15), sauf pour la condition d'agentivité neutre pour arme autonome, sont significatifs. Ces résultats montrent que dans la condition d'agentivité basse, la faute revient au chef, et dans une condition de haute agentivité la faute est attribuée au drone.

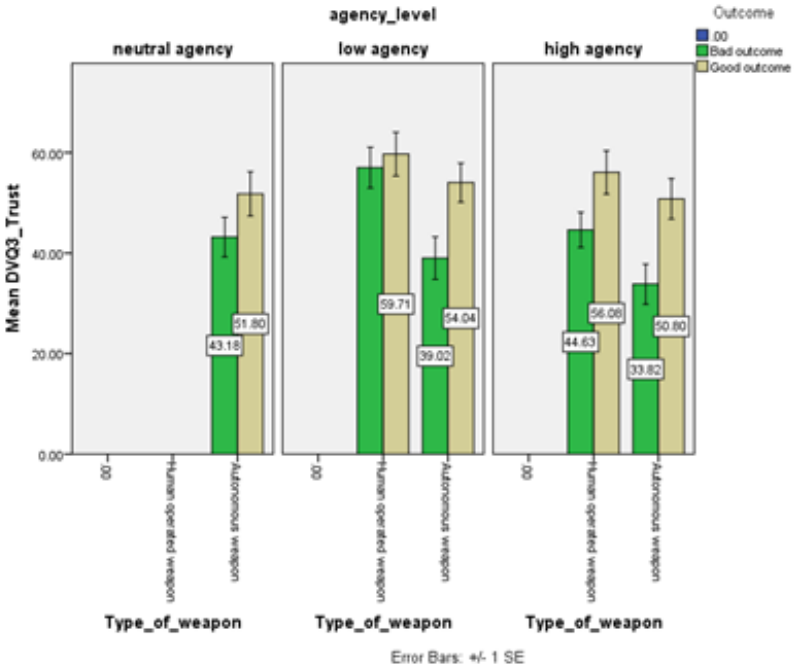


Figure 14 – Moyenne pour valeur «confiance» par condition (de niveau d'agentivité) par résultat

Il est frappant d'observer que la condition à basse agentivité montre une quantité élevée de responsabilité causale pour le

«tort causé». Ce cas de figure est valable pour les drones actionnés par les humains aussi bien que pour les armes autonomes (Figure 16). Contrairement à la variable «faute / blâme», le drone est perçu comme causant beaucoup de mal (tort) aussi bien dans les conditions à haut et bas niveau d'agentivité, mais pour le bas agentivité le tort revient presque de manière égale au chef. Cela signifie que tous deux causent du tort, mais dans le haut niveau, ce tort est beaucoup moins imputé au chef, et le drone est perçu comme étant celui qui cause du tort. Cependant, la différence entre les variables «tort» et «tort causé par le chef» pour le niveau bas d'agentivité de l'arme autonome aussi bien que celle actionnée par l'homme n'est pas significative; par conséquent nous devons être prudents lorsqu'on tire des conclusions sur les effets observés.

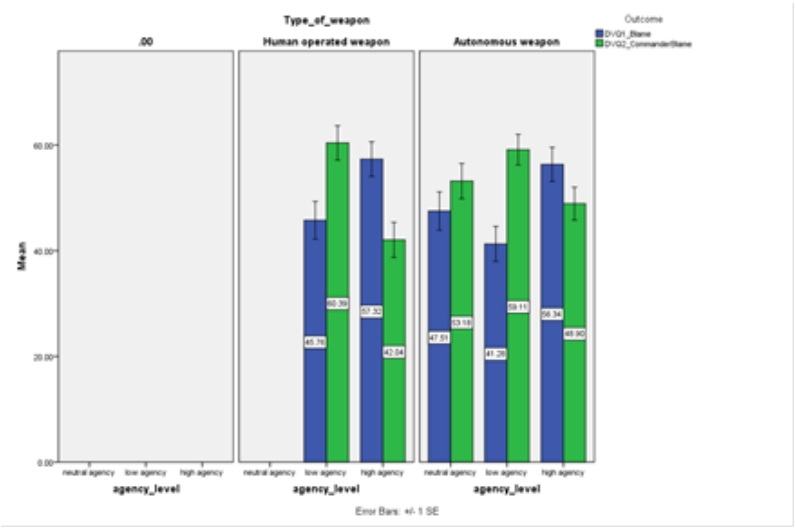


Figure 15 – Moyenne valeurs variables «faute» et «faute du chef» par condition de niveau (d'agentivité) et par type d'arme.

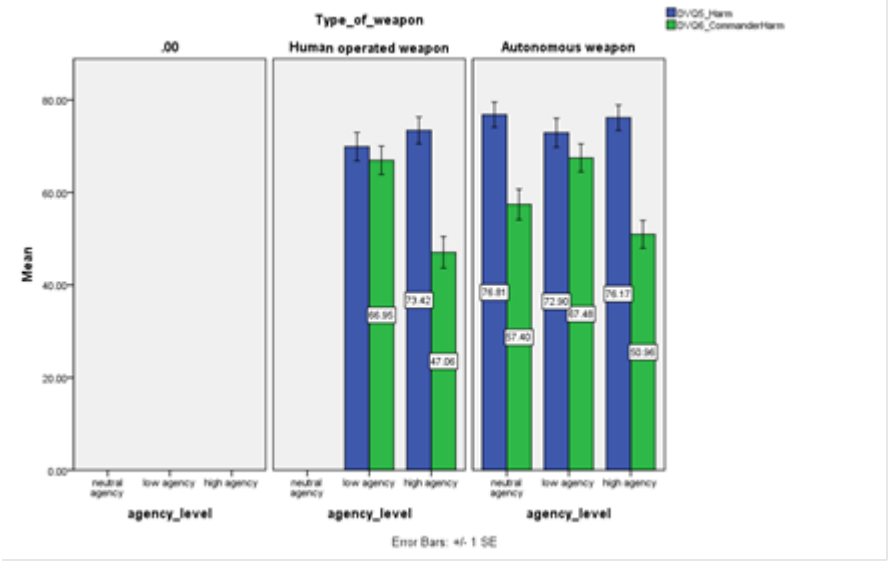


Figure 16 – Moyenne valeurs variables «tort causé» et «tort causé par le chef» par condition de niveau et par type d'arme

4.2.6. Conclusion de l'étude pilote 1

Les résultats de l'étude pilote 1 conduisent aux conclusions suivantes:

- Nous référant aux résultats de l'APC et de l'analyse fiabilité des éléments d'agentivité, nous avons décidé d'enlever SQ3 Règles morales du concept d'agentivité, ce qui améliore la fiabilité du concept: $\alpha = 0.933$. Dans la suite, nous utiliserons le concept agentivité qui contient quatre éléments, *pensée, détermination des objectifs, liberté d'action* et *réalisation des objectifs.*
- Les résultats du Test-T (Student test) indiquent qu'il existe une différence dans la moyenne de la perception de l'agentivité entre les conditions arme autonome et opéra-

tion humaine. Il y a également une grande différence entre les scénarios haut niveau et bas / neutre niveaux d'agentivité. Néanmoins, entre niveaux bas et neutre on n'a pas observé une grande différence. Par conséquent, nous avons décidé de supprimer la condition niveau neutre d'agentivité pour la suite, et de tester seulement les conditions de niveaux bas et élevé.

- L'observation la plus évidente dans la variable *approbation* est qu'il existe une grande différence dans l'approbation entre les situations de bons et de mauvais résultats (Figure 11). Cependant ceci était prévisible au vu de ce qui est noté dans les publications, car James Igoe Walsh (2015) et Kreps (2014) ont constaté que le grand public exprime plus son approbation pour les frappes de drones s'il n'y a pas de dommages collatéraux, comparé aux cas où il y a des dommages collatéraux, mais cela montre que notre manipulation des résultats fonctionne, et nous utiliserons ce point dans la suite.
- Dans l'analyse de la variable *confiance* (Figure 13), nous avons constaté que les personnes sont moins susceptibles de faire confiance à une arme autonome pour prendre les mesures adéquates dans les situations de mauvais résultat qu'à un opérateur humain qui prend des mesures avec le même mauvais résultat. Ceci pourrait révéler une aversion pour les algorithmes qui fait que les algorithmes sont punis plus que les humains qui font des erreurs (Dietvorst, 2016), et nous avons utilisé ces constatations dans l'étude pilote suivante pour étudier plus attentivement l'effet d'aversion pour les algorithmes.
- Nous avons constaté que dans la condition basse agentivité il se trouve un degré étonnamment élevé de responsabilité causale du *tort causé*, et ce phénomène existe également pour les drones actionnés par l'homme et pour les armes

autonomes (Figure 16). Nous avons aussi constaté que le drone est considéré comme causant beaucoup de tort dans les conditions hautes ou basses d'agentivité, mais dans la condition basse le tort est attribué presque à un niveau égal au chef. Bien que ces constatations ne soient pas très significatives, elles sont intéressantes, et méritent d'être approfondies; nous les avons donc conservées pour l'étude pilote 2.

4.3. Etude pilote 2

Dans l'étude pilote 2 nous avons essayé de mieux comprendre l'effet aversion pour l'algorithme; nous avons donc testé des scénarios dans lesquels l'arme autonome pouvait tirer les leçons de ses erreurs (3 + 10), était programmé sur une grande base de données et comprenait sa mission (4 + 11), pouvait être parfois imprévisible (5 + 12), ainsi qu'un scénario qui incluait tous ces aspects (6 + 13). Nous les avons testé dans des cas de bons et de mauvais résultats, et avons ajouté un scénario pour arme autonome avec un bas niveau d'agentivité (1 + 8) pour vérifier si notre manipulation agentivité fonctionnait, ainsi qu'un scénario opérateur humain (7 + 14), dans lequel nous avons posé des questions sur l'opérateur humain et non sur le drone, à la différence de l'étude pilote 1.

Nous avons testé un total de 14 scénarios et avions prévu 50 répondants par scénario, c'est-à-dire 700 réponses au total. Le nombre de répondants par scénario est compris entre 47 et 54. Après le prétraitement des données et avoir éliminé les répondants qui ont échoué au test d'attention, nous avons obtenu 709 réponses complètes.

Tableau11 – Scénarios pour l'étude pilote 2

		Aspects						
		Drone autonome						***HO***
		Niveau 1 – basse agentivité	*Niveau 2 – agentivité élevée + pas d'autre*	*Niveau 3 – agentivité élevée + capacité d'apprendre*	*Niveau 4 – agentivité élevée + maîtrise des procédures*	*Niveau 5 – agentivité élevée + imprévisibilité*	*Niveau 6 – agentivité élevée + tous les aspects*	*Drone actionné par l'homme*
Résultat	***Mauvais (dommages collatéraux)***	*1*	*2*	*3*	*4*	*5*	*6*	*7*
	Bon (pas de dommages collatéraux)	*8*	*9*	*10*	*11*	*12*	*13*	*14*

4.3.1. Analyse fiabilité

Pour les quatre éléments d'agentivité, α = 0.914, ce qui est bien au-dessus du seuil de 0.7 et ne peut être amélioré en éliminant un élément. L'élimination d'autres éléments diminuera la valeur α et réduira la cohérence interne.

Tableau 12 – Résultats analyse fiabilité

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.914	.914	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Thought	208.4827	6427.511	.801	.647	.890
Goal_setting	204.9885	6499.803	.834	.695	.877
Act_freely	203.4020	6782.379	.804	.646	.888
Achieve_goals	199.4975	7114.019	.781	.614	.897

4.3.2. Analyse composante principale (ACP)

L'ACP montre que la *réflexion, détermination des objectifs, liberté d'action* et *réalisation des objectifs* peuvent être considérées comme un seul ensemble qui donne 79.60 % de variation dans les variables originales (Tableau 13). Ceci se constate aussi dans le diagramme d'éboulis qui montre les valeurs propres des composantes (Figure 17). L'ACP montre que les quatre éléments *réflexion, détermination des objectifs, liberté d'action* et *réalisation des objec-*

tifs peuvent être considérées comme les composantes qui sous-tendent le concept d'agentivité (Tableau 13).

Tableau 13 – Résultats ACP étude pilote 2

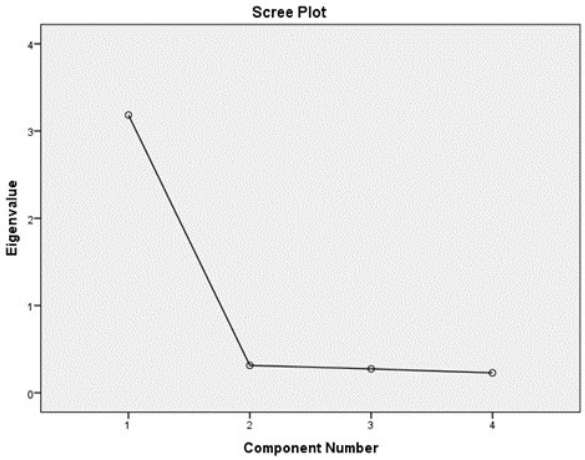


Figure 17 – Diagramme d'éboulis ACP étude pilote 2

Component Matrix[a]

	Component	
	1	2
Goal_setting	.910	
Act_freely	.892	
Thought	.889	
Achieve_goals	.877	.446

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

4.3.3. Analyse de corrélation

La corrélation entre le concept d'agentivité et les variables dépendantes n'est pas grande, mais significative pour tous les ensembles ($p < .05$) (Tableau 14). Les mêmes effets concernant les variables «tort causé» et «faute / blâme» comme pour l'étude pilote 1 peuvent être observés ici. Il existe une relation négative entre le concept d'agentivité et le niveau d'inquiétude: lorsque la perception de l'agentivité augmente, les personnes indiquent qu'elles ressentent plus d'anxiété, mais cet effet est de faible amplitude ($r = -.076$). L'analyse de corrélation n'est pas assez détaillée pour s'attarder sur ces résultats; par conséquent ceci sera fait dans l'analyse des dépendantes variables.

4.3.4. Vérification manipulation sur l'agentivité

Le graphique de la Figure 18 montre qu'il y a une grande différence entre les scénarios de basse et de haute agentivité, et cette différence est significative. La différence confirme que notre manipulation sur l'agentivité fonctionne. L'agentivité pour l'opérateur humain et les drones d'agentivité élevée sont au même niveau (dans la présente étude nous avons posé des questions spécifiques sur l'opérateur humain, alors que c'était sur le drone dans la précédente enquête). Il y a une différence entre les conditions bons résultats et mauvais résultats, et bien que ces effets soient plus sensibles pour les conditions dans lesquelles nous avons porté notre attention sur l'aversion envers l'algorithme, la différence n'est pas très grande (en moyenne de 8 points sur une échelle de 100). Également, les barres d'erreur entre le bon résultat et le mauvais résultat pour tous les scénarios se chevauchent; ceci indique que ce ne sont pas des ensembles distincts. La seule différence évidente et significative dans des ensembles ($p < .05$) se situe entre la condition agentivité basse arme autonome et les autres scénarios. Ceci est également confirmé par le niveau d'importance des

tests-T d'échantillons indépendants, que nous avons décidé de ne pas présenter sur le graphique ci-dessous par souci de clarté.

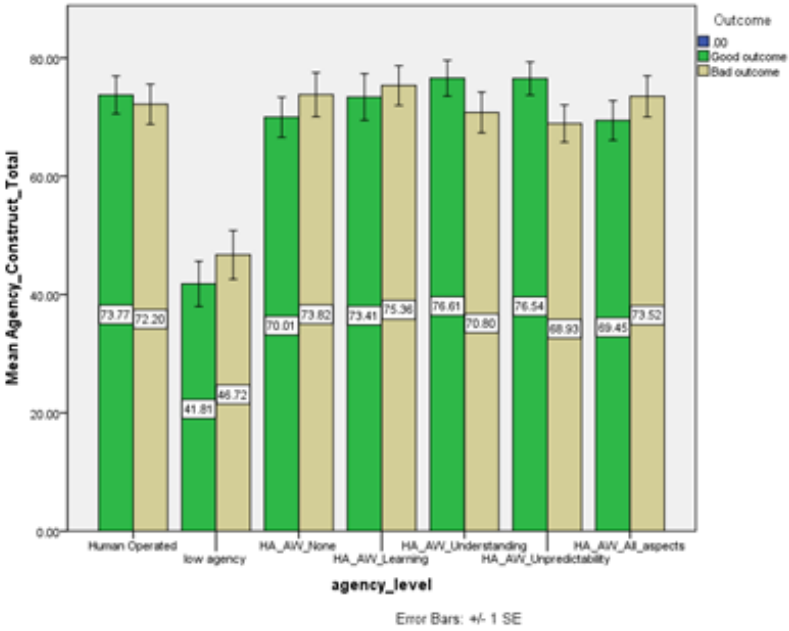


Figure 18 – Valeur moyenne concept agentivité par condition par résultat

4.3.5. Analyse des variables dépendantes

La corrélation entre le concept d'agentivité et les variables dépendantes est partout significative ($p < .05$); donc les constatations les plus frappantes pour chacune des variables dépendantes sont décrites ci-dessous.

Faute / blâme

La constatation la plus remarquable pour «faute» est que les

Correlations

		Agency_Construct_Total	DVQ1_Blame	DVQ2_CommanderBlame	DVQ3_Trust	DVQ4_CommanderTrust	DVQ5_Harm	DVQ6_CommanderHarm	DVQ7_Dignity	DVQ8_Confidence	DVQ9_Expectations	DVQ10_Support	DVQ11_Fair	DVQ12_Uneasy
Agency_Construct_Total	Pearson Correlation	1	.271**	-.267**	.205**	.217**	.124**	-.198**	.102**	.209**	.146**	.164**	.142**	-.076*
	Sig. (2-tailed)		.000	.000	.000	.000	.001	.000	.007	.000	.000	.000	.000	.043
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ1_Blame	Pearson Correlation	.271**	1	-.069	-.028	-.046	.258**	-.023	-.056	-.048	-.137**	-.118**	-.092*	.180**
	Sig. (2-tailed)	.000		.065	.464	.226	.000	.534	.138	.198	.000	.002	.014	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ2_CommanderBlame	Pearson Correlation	-.267**	-.069	1	-.287**	-.278**	.152**	.691**	-.206**	-.300**	-.205**	-.250**	-.275**	.312**
	Sig. (2-tailed)	.000	.065		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ3_Trust	Pearson Correlation	.205**	-.028	-.287**	1	.653**	-.281**	-.224**	.595**	.901**	.533**	.726**	.722**	-.616**
	Sig. (2-tailed)	.000	.464	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ4_CommanderTrust	Pearson Correlation	.217**	-.046	-.278**	.653**	1	-.185**	-.228**	.432**	.635**	.425**	.631**	.552**	-.458**
	Sig. (2-tailed)	.000	.226	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ5_Harm	Pearson Correlation	.124**	.258**	.152**	-.281**	-.185**	1	.311**	-.374**	-.301**	-.217**	-.253**	-.324**	.395**
	Sig. (2-tailed)	.001	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ6_CommanderHarm	Pearson Correlation	-.198**	-.023	.691**	-.224**	-.228**	.311**	1	-.183**	-.226**	-.203**	-.235**	-.248**	.289**
	Sig. (2-tailed)	.000	.534	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ7_Dignity	Pearson Correlation	.102**	-.056	-.206**	.595**	.432**	-.374**	-.183**	1	.588**	.415**	.533**	.598**	-.509**
	Sig. (2-tailed)	.007	.138	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ8_Confidence	Pearson Correlation	.209**	-.048	-.300**	.901**	.635**	-.301**	-.226**	.588**	1	.553**	.723**	.725**	-.649**
	Sig. (2-tailed)	.000	.198	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ9_Expectations	Pearson Correlation	.146**	-.137**	-.205**	.533**	.425**	-.217**	-.203**	.415**	.553**	1	.511**	.621**	-.509**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ10_Support	Pearson Correlation	.164**	-.118**	-.250**	.726**	.631**	-.253**	-.235**	.533**	.723**	.511**	1	.685**	-.651**
	Sig. (2-tailed)	.000	.002	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ11_Fair	Pearson Correlation	.142**	-.092*	-.275**	.722**	.552**	-.324**	-.248**	.598**	.725**	.621**	.685**	1	-.656**
	Sig. (2-tailed)	.000	.014	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
	N	708	708	708	708	708	708	708	708	708	708	708	708	708
DVQ12_Uneasy	Pearson Correlation	-.076*	.180**	.312**	-.616**	-.458**	.395**	.289**	-.509**	-.649**	-.509**	-.651**	-.656**	1
	Sig. (2-tailed)	.043	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	708	708	708	708	708	708	708	708	708	708	708	708	708

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Tableau 14 – Matrice de corrélation entre concept agentivité et variables dépendantes étude pilote 2

personnes attribuent moins de blâme dans la condition de basse agentivité si l'on compare avec les conditions d'agentivité élevée pour les armes autonomes et opérateur humain. Pour ces conditions, la différence est importante ($p < .05$). Quelques effets de l'algorithme peuvent être constatés dans le scénario d'apprentissage et de compréhension, où les personnes attribuent moins de blâme lorsque l'arme autonome ne fait pas d'erreur, et plus de blâme lorsqu'elle fait une erreur avec un mauvais résultat. Cependant, ces conditions montrent un chevauchement des barres d'erreur, et ne sont pas significatives.

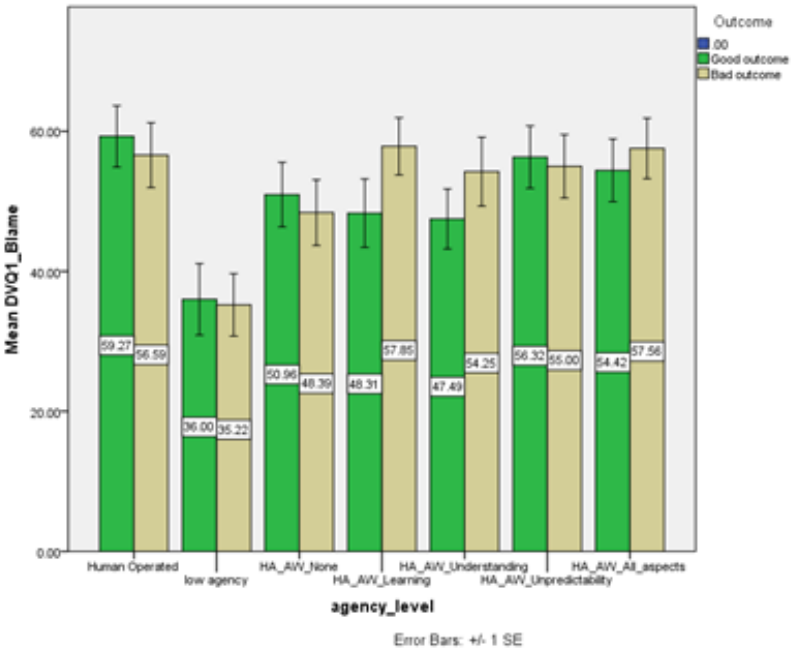


Figure 19 – Valeur moyenne variable «blâme» par condition par résultat

Faute du chef

L'observation la plus intéressante dans la variable *faute du chef (blâme)* et que la faute, est attribuée au chef beaucoup moins lorsque l'arme autonome fait une erreur dans les conditions «tirer les leçons» et «imprévisibilité», et ceci est significatif ($p < .05$). Il pourrait arriver que la faute soit attribuée au drone puisque les personnes s'attendent à ce qu'il soit capable de tirer les leçons de ses erreurs ou qu'il se conduise de manière imprévisible, et qu'il ne faut pas accuser *le chef*. Également, dans la condition basse agentivité le *chef est accusé* un peu plus que dans les autres conditions.

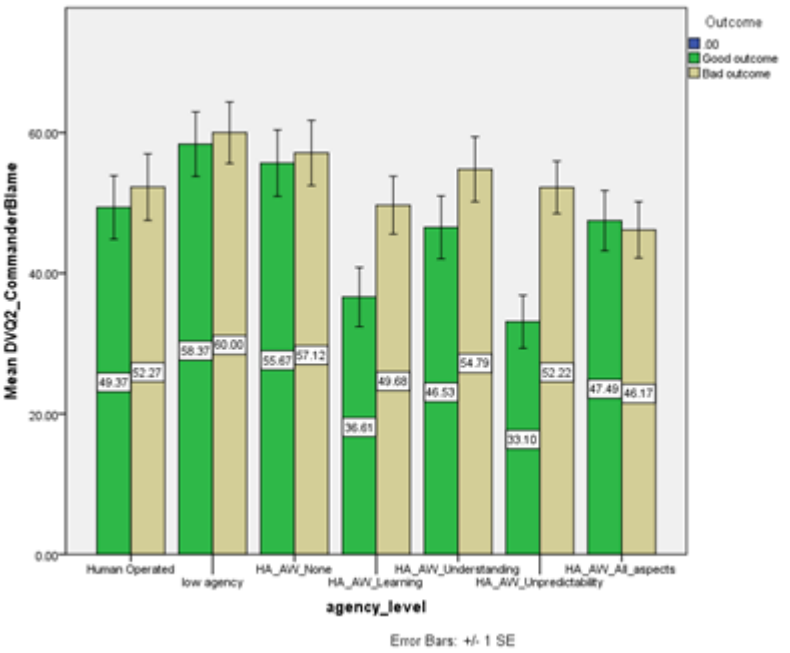


Figure 20 – Valeur moyenne variable «faute du chef» par condition par résultat

Confiance

Quelle que soit la condition on peut constater que les personnes font davantage confiance à l'arme autonome ou à l'opérateur humain pour prendre les mesures adéquates après un bon résultat qu'après un mauvais résultat. Sauf dans la condition de basse agentivité, la différence de confiance dans le cas de bon résultat et dans celui de mauvais résultat est significative ($p < .05$). Les répondants font confiance à l'opérateur humain plus qu'à l'arme autonome, en particulier dans la condition de basse agentivité, et la différence est significative ($p < .05$).

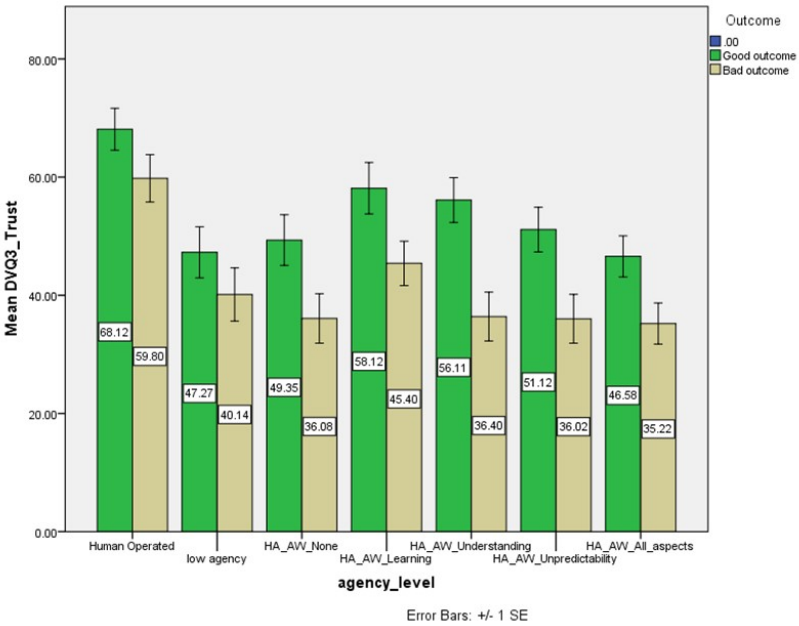


Figure 21 – Valeur moyenne variable «confiance» par condition par résultat

Confiance en le chef

Globalement les répondants font confiance au chef pour prendre les mesures adéquates dans les conditions «bon résultat», en comparaison des conditions «mauvais résultat». La confiance accordée au chef diffère dans les conditions de basse agentivité et de haute agentivité, sans autre information, et la différence est significative ($p < .05$). Savoir que l'arme autonome a des capacités de compréhension entraîne un plus haut degré de confiance dans les conditions «mauvais résultat», et ceci est significatif ($p < .05$).

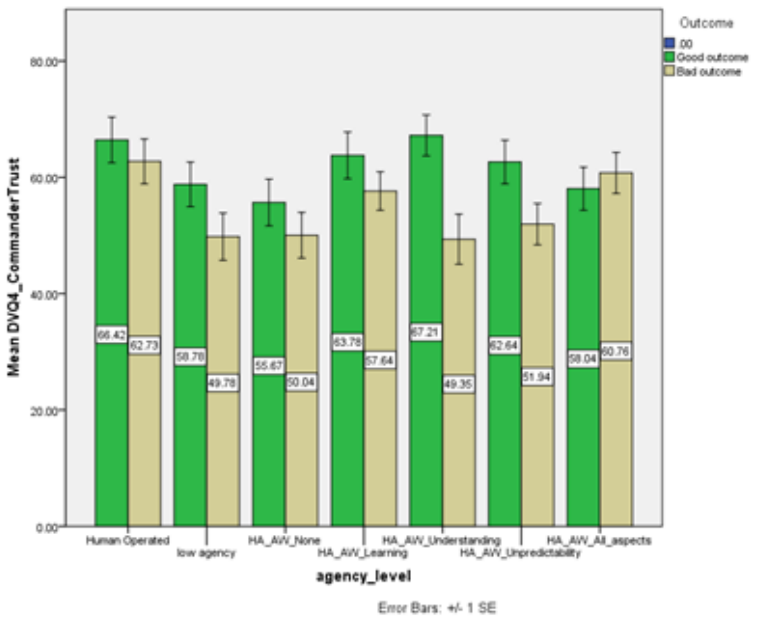


Figure 22 – valeur moyenne variable «confiance en le chef» par condition par résultat

Tort causé

Dans les conditions «mauvais résultat», les répondants perçoivent plus de tort causé que dans les conditions «bons résultats», ce qui n'est pas étonnant, et la différence est significative dans presque toutes les conditions ($p < .05$). Le plus intéressant dans ce graphique est que lorsque l'arme autonome se conduit de manière imprévisible, le type de résultat paraît ne pas compter, et la perception du «tort causé» est la même; mais la différence entre les deux ensembles «bons ou mauvais résultats» n'est pas significative.

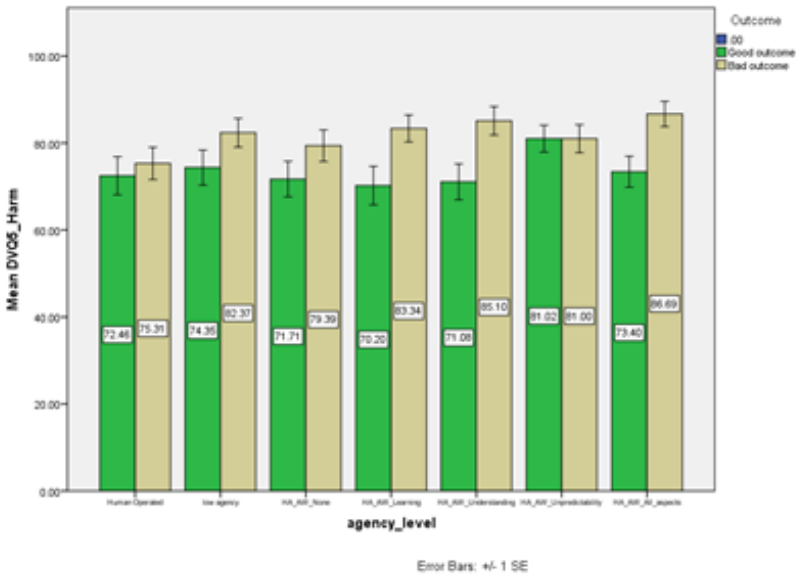


Figure 23 – valeur moyenne variable «tort causé» par condition par résultat

Tort causé par le chef

Dans les conditions «mauvais résultat» les répondants perçoivent que le chef cause plus de tort que dans les conditions «bon résultat», ce qui n'est pas étonnant. Plus intéressante est la grande différence dans la condition «apprentissage» entre le bon et le mauvais résultat où on attribue au chef beaucoup moins de tort dans la condition «bon résultat», ce qui est significatif ($p < .05$). La différence entre les conditions «bon» et «mauvais» résultats dans les scénarios opérateur humain et imprévisibilité agentivité élevée est également significative ($p < .05$).

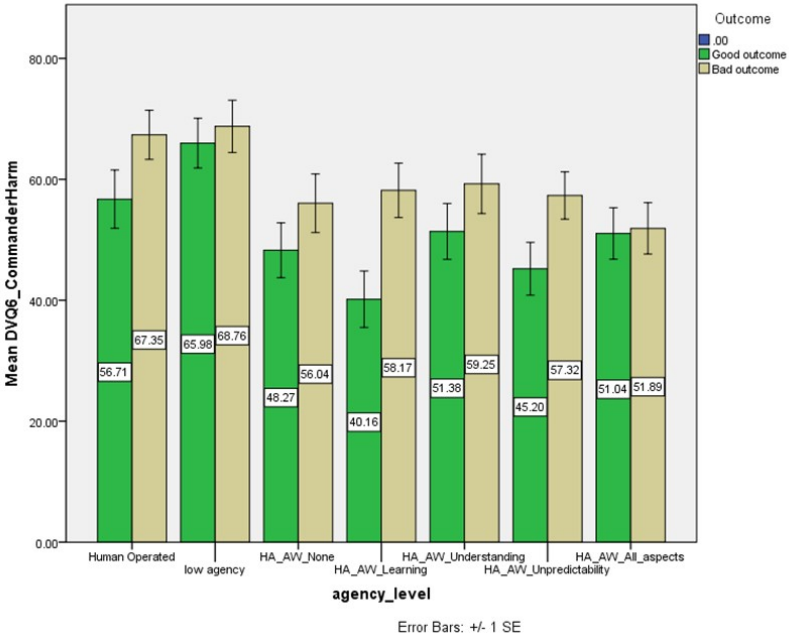


Figure 24 – valeur moyenne variable «tort causé par le chef» par condition par résultat

Dignité humaine

L'opérateur humain agit avec beaucoup plus de respect envers la dignité humaine que l'arme autonome, sauf dans les conditions dans lesquelles l'arme autonome est programmée sur un large éventail de données pour sa mission, et cette différence est significative ($p < .05$). Également, le fait que l'arme autonome fasse une erreur est important, car nous observons des différences importantes ($p < .05$) entre les scénarios «bon» et «mauvais» résultats pour les scénarios armes autonomes en agentivité basse, aucun des aspects en agentivité élevée et tous les aspects en agentivité haute.

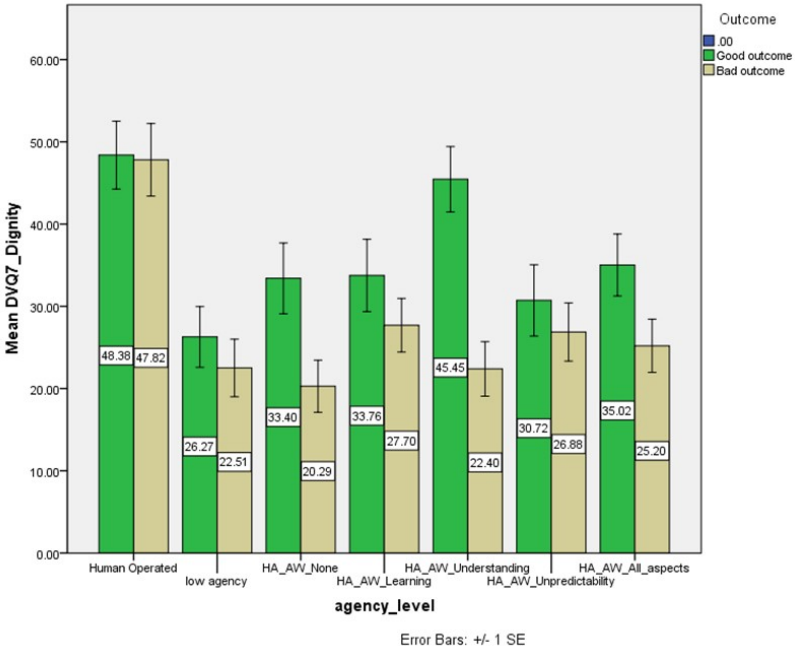


Figure 25 – valeur moyenne – variable «dignité humaine» par condition par résultat

Assurance

Globalement les répondants sont davantage assurés que l'arme autonome prendra les mesures adéquates après un bon résultat qu'après un mauvais résultat, et toutes les différences sont significatives ($p < .05$). Ils ont également un peu plus confiance en l'opérateur humain qu'en l'arme autonome. Les différences entre ces scénarios sont significatives ($p < .05$). Lorsque l'arme autonome a des capacités d'apprentissage, leur assurance est un peu plus haute si l'on compare aux autres conditions pour les armes autonomes.

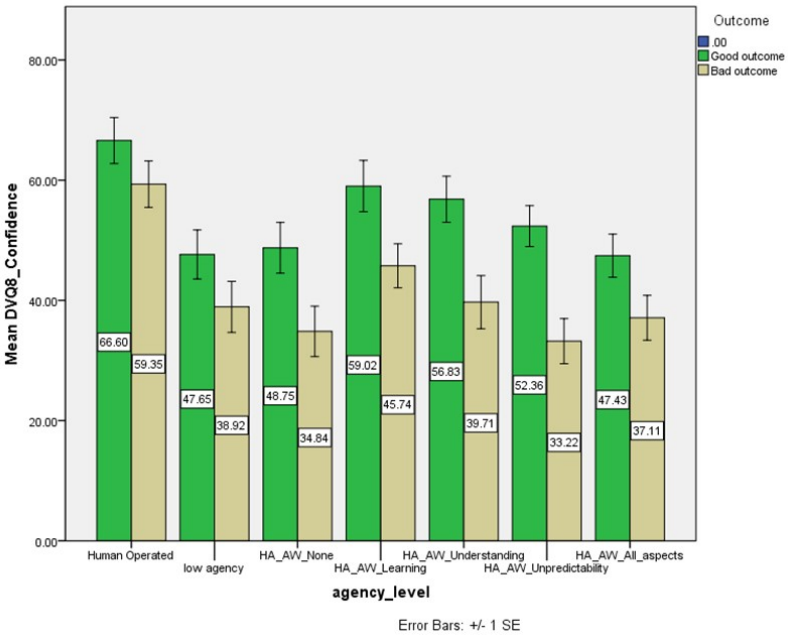


Figure 26 – valeur moyenne variable «assurance» par condition par résultat

Attentes

Dans les scénarios «bons résultats», l'opérateur humain et l'arme autonome agissent tous deux davantage dans le sens des attentes des répondants que dans les scénarios «mauvais résultats». Toutes les différences entre les conditions «bons résultats» et «mauvais résultats» sont significatives ($p < .05$). Les attentes sont plus basses dans la condition basse agentivité lorsqu'il y a mauvais résultat, mais il n'y a pas beaucoup de différence si l'on compare avec les autres conditions.

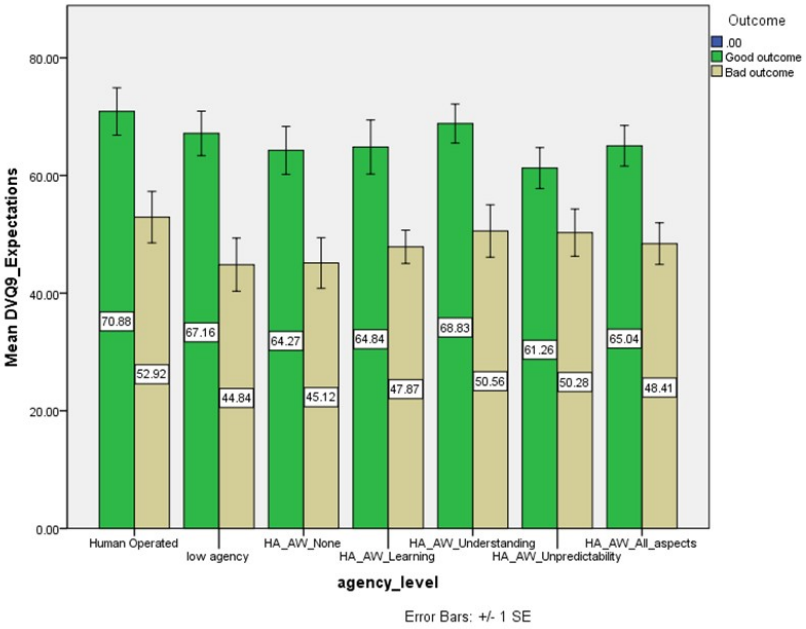


Figure 27 – valeur moyenne variable «attentes» par condition par résultat

Approbation / soutien

Les répondants manifestent plus d'approbation envers les drones actionnés par les humains que pour les armes auto-

nomes. L'approbation est la moins élevée, mais de manière non significative, pour les conditions de moindre agentivité si l'on compare avec les scénarios dans lesquels les informations sur l'arme autonome sont fournies. Également, l'approbation est plus haute pour les scénarios avec «bon» résultat que pour ceux ayant un «mauvais» résultat; sauf pour la condition «activé par l'homme» et «arme autonome basse agentivité», ces différences sont significatives ($p < .05$).

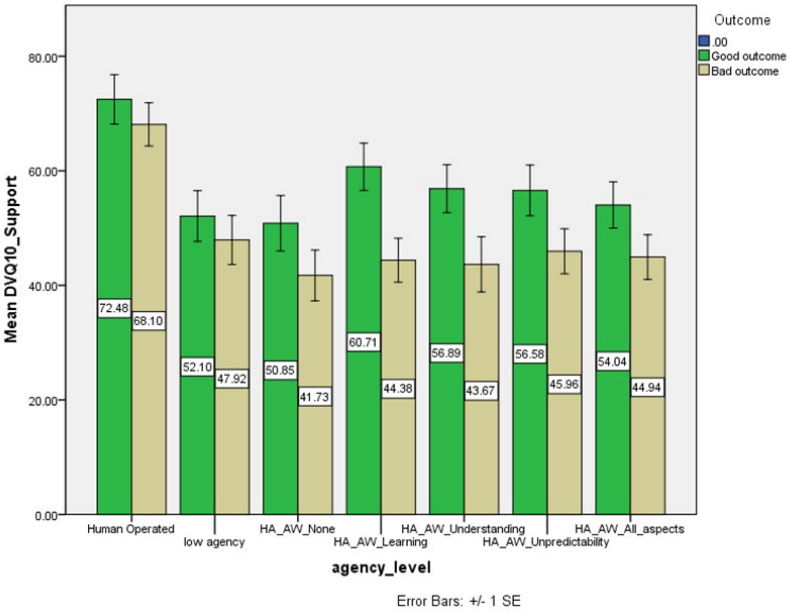


Figure 28 – valeur moyenne variable «approbation» par condition par résultat

Equité

Globalement, les actions des opérateurs humains sont considérées comme étant plus équitables que celles des armes autonomes. La perception de l'équité est beaucoup plus haute dans

les scénarios à bons résultats que dans les scénarios à mauvais résultats, et dans tous les scénarios ces différences sont significatives ($p < .05$). La différence entre la condition d'agentivité des armes autonomes n'est pas importante (< 10 points sur une échelle de 100).

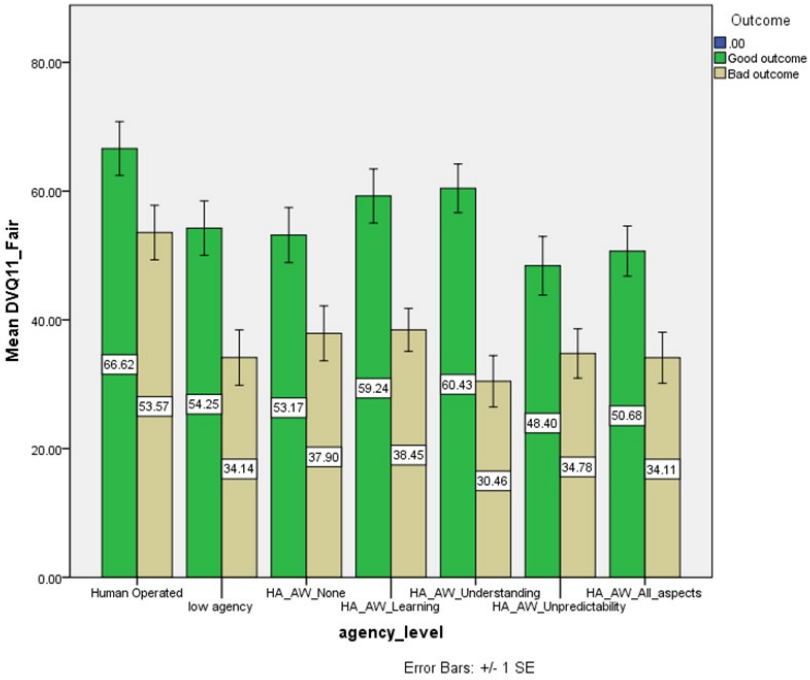


Figure 29 – valeur moyenne variable «équité» par condition par résultat

Inquiétude / anxiété

Dans les scénarios à mauvais résultats les répondants ressentent plus d'inquiétude concernant les actions de l'opérateur humain aussi bien que celles de l'arme autonome. Il n'y a pas grande différence entre les niveaux d'inquiétude pour les différentes conditions des armes autonomes. Le niveau d'inquiétude dans le cas de l'opérateur humain est beaucoup plus bas que

celui ayant trait à l'arme autonome. Sauf pour la condition «agentivité élevée pas d'autres aspects armes autonomes», toutes les différences entre les scénarios bons et mauvais résultats sont significatives ($p < .05$).

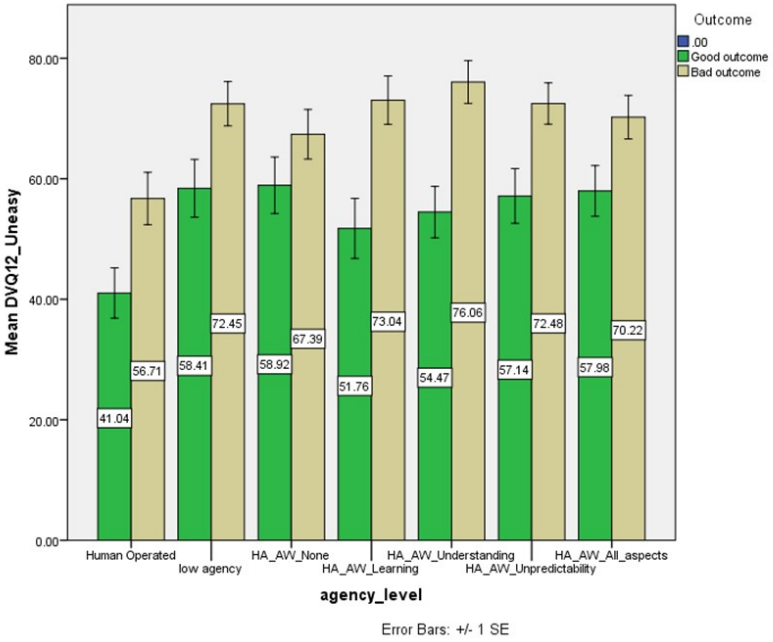


Figure 30 – valeur moyenne variable «inquiétude» par condition par résultat

4.3.6. Conclusion étude pilote 2

Les résultats de l'étude pilote 2 conduisent aux conclusions suivantes:

- Le test fiabilité a démontré que les quatre éléments du concept d'agentivité sont fiables, et selon l'ACP peuvent être considérés comme une seule composante;

- La corrélation entre le coefficient du concept agentivité et toutes les variables dépendantes est significative et vont dans la même direction que pour l'étude pilote 1. Nous avons constaté un rapport négatif entre le concept agentivité et le niveau d'inquiétude, ce qui signifie que lorsque la perception de l'agentivité augmente, les répondants indiquent qu'ils ressentent plus d'inquiétude, mais cet effet reste très faible et approche de 0 ($r = -.076$). Afin de voir si nous pouvions obtenir un effet plus important, nous avons changé la formulation de la question de la variable «inquiétude» dans l'étude finale;
- La manipulation de l'agentivité montre qu'il y a une grande différence entre les scénarios haute agentivité et le scénario basse agentivité. L'agentivité de l'opérateur humain et la haute agentivité des armes autonomes sont au même niveau. Cependant, les moyennes du concept d'agentivité pour les scénarios à haute agentivité qui incluent l'effet aversion envers l'algorithme sont presque identiques. Pour préciser l'effet aversion envers l'algorithme il nous faut étudier davantage les publications existantes et effectuer plus d'études pilotes. Du fait des contraintes de temps imparti, nous n'avons pas poursuivi cette démarche, et avons décidé de concentrer l'étude finale sur la perception de l'agentivité.
- Un grand nombre des effets que nous avons constatés pour les variables dépendantes devaient, comme nous nous y attendions, être basés sur les résultats de l'étude pilote1. Nous n'avons pas observé d'effets distincts dans les VD entre les différentes conditions d'agentivité des armes autonomes. Nous avons vérifié pour voir si nous pouvions calquer l'effet aversion pour l'algorithme de l'étude pilote1, dans laquelle le «tort causé» passait du drone au chef dans la condition de basse agentivité, mais malheureusement cet

effet n'a pas pu être observé dans l'étude pilote2. Même s'il serait bon d'étudier d'un peu plus près cette chaîne de responsabilité, nous avons décidé de ne pas l'inclure dans l'étude finale, puisque nous n'étions pas vraiment sûrs de ce qui se passe; il faudrait étudier de plus près les publications et procéder à d'autres études pilote.

4.4. Etude finale – échantillon militaire

Afin de déterminer le nombre de scénarios pour l'étude finale, nous avons procédé à des calculs Power pour évaluer le nombre total de participants dont nous aurions besoin en nous basant sur les résultats de la première étude pilote, car nous utiliserons ces scénarios également pour l'étude finale. En nous fondant sur des calculs de puissance (ampleur d'effet 0.4, ampleur statistique souhaitée de 0.8 et niveau de probabilité de 0.05), nous avons compté sur un total de 200 réponses, et déterminé que nous pouvions donner 3 scénarios. Nous avons décidé de privilégier la perception de l'agentivité et de comparer la technologie actuelle, un drone actionné par l'homme, avec la technologie du futur, c'est-à-dire l'arme autonome à haute agentivité (Tableau 15). Nous avons également ajouté un scénario avec arme autonome à agentivité neutre, pour comparer la perception de l'agentivité dans les trois conditions. Dans cette étude, nous avons posé des questions spécifiques sur le drone et non sur l'opérateur humain qui le contrôle pour comparer avec la perception de l'agentivité des artefacts. Comme nous ne pouvions travailler que sur trois scénarios, nous avons privilégié la condition «mauvais résultat», car les études pilotes montraient les effets les plus distincts de cette condition.

Nous avons diffusé l'enquête par email avec un lien anonyme et donc n'avions pas de garantie quant au nombre des répondants. Après le prétraitement des données et après avoir éliminé les répondants qui avaient échoué au test d'attention, il

restait 239 réponses pour l'échantillon complet. Le nombre de répondants par scénario s'élevait entre 64 et 96, ce qui est curieux car les répondants auraient dû être également répartis entre les scénarios. L'échantillon de 239 réponses est un mélange de personnels militaires et civils néerlandais travaillant pour le Ministère de la Défense, car si nous avions exclu les civils seuls restaient 149 répondants, ce qui était trop réduit pour une analyse de données crédible.

Tableau 15 – scénarios étude finale

	Niveaux d'agentivité		
	Actionné par l'homme	Agentivité neutre arme autonome	Agentivité élevée arme autonome
Mauvais résultat	1	2	3

Pour ce tableau, les définitions suivantes sont utilisées:

- *Drone actionné par l'homme:* drone directement contrôlé par un être humain. Par hypothèse, perçu comme ayant une très basse agentivité.
- *Arme autonome à haute agentivité:* on dit aux participants que l'arme autonome possède des caractéristiques d'agentivité (comme peser le pour et le contre et faire des choix de manière indépendante). Par hypothèse, perçu comme possédant une très haute agentivité.
- *Arme autonome, agentivité neutre:* aucune information au sujet des caractéristiques d'agentivité de l'arme autonome n'est indiquée. La force de cette approche est que nous pouvons recueillir des données sur la manière dont les participants utilisent leurs idées préconçues sur les armes autonomes pour comprendre les actions de l'arme autonome, et com-

parer ces idées aux conditions drone actionné par l'homme et arme autonome à haute agentivité.

4.4.1. Analyse fiabilité

Pour les quatre éléments agentivité α = .865, ce qui est au-dessus du seuil de 0.7 et ne peut être amélioré en éliminant un élément, ce qui signifie que le concept agentivité est fiable (Tableau 16). L'élimination d'autres éléments diminue la valeur α et réduit la cohérence interne.

Tableau 16 – Résultats analyse fiabilité des éléments d'agentivité

Reliability Statistics

Cronbach's Alpha	N of Items
.865	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Thought	116.43	7175.573	.620	.863
Goal setting	123.33	6342.474	.756	.810
Act freely	122.62	6566.723	.740	.816
Achieve goals	118.84	6459.146	.740	.816

4.4.2. Analyse composante principale (ACP)

L'APC montre que la *Réflexion, la détermination des objectifs, la liberté d'action* et *la réalisation des objectifs* peuvent être considérés comme un seul ensemble qui constitue 71.18 % des variations dans les variables d'origine. Ceci peut également être constaté

dans le diagramme d'éboulis (scree plot) qui montre les valeurs propres des composantes (Figure 31). L'APC montre que les quatre éléments *réflexion, détermination des objectifs, liberté d'action* et *réalisation des objectifs* peuvent être considérés comme une seule composante du concept d'agentivité (Tableau 17).

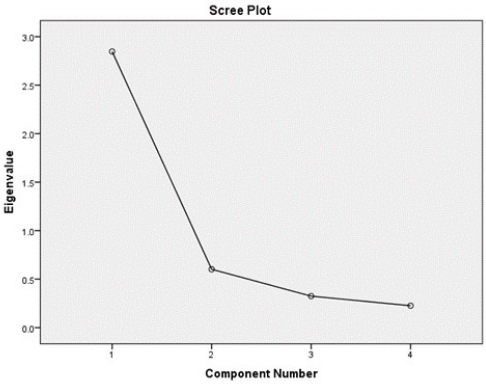


Figure 31 – diagramme d'éboulis APC étude finale

Component Matrix[a]

	Component	
	1	2
Thought	.774	.560
Goal setting	.873	
Act freely	.861	
Achieve goals	.863	

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Tableau 17 – résultats éléments d'agentivité étude finale

4.4.3. Analyse de corrélation concept d'agentivité

La corrélation entre le concept d'agentivité et les variables dépendantes est plus grande que dans les études pilotes précé-

dentes, mais significative pour seulement 7 des 9 variables; *confiance, dignité humaine, assurance, attentes, approbation, équité et inquiétude* ($p < .01$) (Tableau 18). Il existe un rapport négatif entre le concept d'agentivité et le degré d'inquiétude qui est plus haut que précédemment dans les études pilotes. Ceci indique que lorsque la perception de l'agentivité augmente, les personnes déclarent qu'elles sont plus inquiètes. L'analyse de corrélation n'est pas assez détaillée pour étudier de près ces résultats, et donc cela sera fait dans l'analyse des variables dépendantes.

Correlations

		Agency	DVQ1_Blame	DVQ2_Trust	DVQ3_Harm	DVQ4_Human_Dignity	DVQ5_Confidence	DVQ6_Expectations	DVQ7_Support	DVQ8_Fair	DVQ9_Anxiety
Agency	Pearson Correlation	1	.009	.406**	-.070	.347**	.428**	.320**	.337**	.451**	-.216**
	Sig. (2-tailed)		.888	.000	.275	.000	.000	.000	.000	.000	.001
	N	248	248	248	248	248	248	248	248	248	248
DVQ1_Blame	Pearson Correlation	.009	1	-.039	.056	.018	.027	-.109	-.002	-.080	.128*
	Sig. (2-tailed)	.888		.545	.379	.783	.667	.086	.979	.209	.044
	N	248	248	248	248	248	248	248	248	248	248
DVQ2_Trust	Pearson Correlation	.406**	-.039	1	-.217**	.482**	.767**	.324**	.597**	.540**	-.451**
	Sig. (2-tailed)	.000	.545		.001	.000	.000	.000	.000	.000	.000
	N	248	248	248	248	248	248	248	248	248	248
DVQ3_Harm	Pearson Correlation	-.070	.056	-.217**	1	-.180**	-.197**	-.155*	-.133*	-.200**	.360**
	Sig. (2-tailed)	.275	.379	.001		.004	.002	.015	.036	.002	.000
	N	248	248	248	248	248	248	248	248	248	248
DVQ4_Human_Dignity	Pearson Correlation	.347**	.018	.482**	-.180**	1	.564**	.239**	.564**	.472**	-.385**
	Sig. (2-tailed)	.000	.783	.000	.004		.000	.000	.000	.000	.000
	N	248	248	248	248	248	248	248	248	248	248
DVQ5_Confidence	Pearson Correlation	.428**	.027	.767**	-.197**	.564**	1	.376**	.685**	.600**	-.495**
	Sig. (2-tailed)	.000	.667	.000	.002	.000		.000	.000	.000	.000
	N	248	248	248	248	248	248	248	248	248	248
DVQ6_Expectations	Pearson Correlation	.320**	-.109	.324**	-.155*	.239**	.376**	1	.271**	.475**	-.264**
	Sig. (2-tailed)	.000	.086	.000	.015	.000	.000		.000	.000	.000
	N	248	248	248	248	248	248	248	248	248	248
DVQ7_Support	Pearson Correlation	.337**	-.002	.597**	-.133*	.564**	.685**	.271**	1	.528**	-.488**
	Sig. (2-tailed)	.000	.979	.000	.036	.000	.000	.000		.000	.000
	N	248	248	248	248	248	248	248	248	248	248
DVQ8_Fair	Pearson Correlation	.451**	-.080	.540**	-.200**	.472**	.600**	.475**	.528**	1	-.440**
	Sig. (2-tailed)	.000	.209	.000	.002	.000	.000	.000	.000		.000
	N	248	248	248	248	248	248	248	248	248	248
DVQ9_Anxiety	Pearson Correlation	-.216**	.128*	-.451**	.360**	-.385**	-.495**	-.264**	-.488**	-.440**	1
	Sig. (2-tailed)	.001	.044	.000	.000	.000	.000	.000	.000	.000	
	N	248	248	248	248	248	248	248	248	248	248

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Tableau 18 – Corrélations entre le concept agentivité et les variables dépendantes pour l'étude finale

4.4.4. Vérification manipulation sur l'agentivité

Le graphique de la Figure 32 montre une augmentation de la perception de l'agentivité sur toutes les conditions. La différence dans la perception de l'agentivité entre le scénario «arme actionné par l'homme» et le scénario «arme autonome agentivité neutre» est significative, donc ces ensembles peuvent être considérés comme différents. Le test de Student (T-test) sur les échantillons indépendants pour le scénario «arme autonome agentivité neutre» et les scénarios «arme autonome agentivité élevée» donne $p = .063$, juste au-dessus du seuil de pertinence. Le T-test indépendant pour la condition «arme actionnée par l'homme» et la condition «arme autonome agentivité neutre» montre que $p = .001$, et pour les conditions «arme actionnée par l'homme» et «arme autonome agentivité élevée», $p = .000$.

Nous avons posé comme hypothèse que le drone actionné par l'homme sera perçu comme ayant une agentivité basse, et l'arme autonome à haute agentivité sera perçue comme ayant une agentivité élevée, et les résultats montrent que c'est bien le cas. L'analyse centrale s'intéressait à la manière dont l'agentivité neutre peut être comparée aux deux autres cas. Nous avons supposé que les personnels militaires ne percevraient pas les armes à agentivité neutre comme ayant des états mentaux. Par conséquent, nous nous attendions à aucune différence entre la condition «arme autonome à agentivité neutre» et la condition dans laquelle un humain actionne un drone à distance. Nous nous attendions aussi à ce que la condition «agentivité neutre» soit considérée comme étant significativement différente de la condition «haute agentivité».

Les résultats confirment que le drone actionné par l'homme est perçu comme ayant une agentivité basse, et que l'arme autonome est perçue comme ayant une agentivité élevée. Cependant, les résultats montrent qu'une arme autonome à agentivité neutre est perçue comme ayant plus d'agentivité

qu'un drone opéré par l'homme, et que la perception de l'agentivité dans la condition «haute agentivité» est juste légèrement plus haute. Cela signifie que nous devons rejeter notre hypothèse, et conclure que les militaires et les civils travaillant pour le Ministère de la Défense néerlandais perçoivent les armes autonomes comme ayant des propriétés d'agentivité.

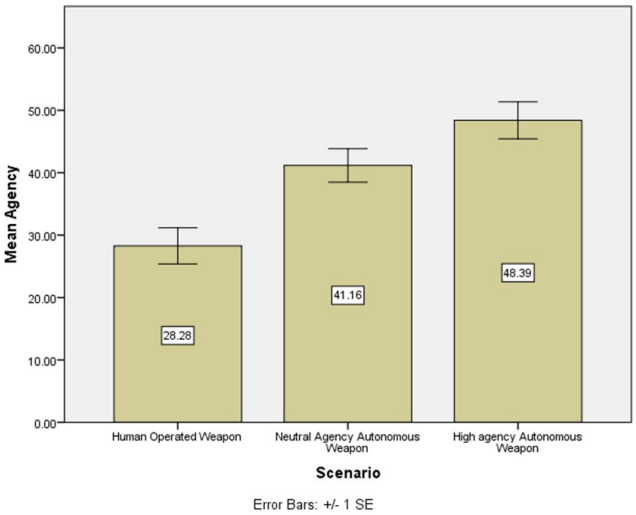


Figure 32 – valeur moyenne d'agentivité par condition

Dans la Figure 33, si nous considérons la différence entre les répondants militaires et civils, nous constatons quelque chose de très intéressant, mais comme les deux groupes ne sont pas significativement différents dans deux des scénarios, seulement dans la condition «agentivité élevée», il nous faut être prudents dans l'interprétation des résultats; l'observation qui suit n'est donc donnée qu'à titre d'illustration. Dans le scénario «arme actionnée par l'homme» ainsi que dans le scénario «agentivité neutre», la perception de l'agentivité des répondants militaires et civils est au même niveau. Cependant, dans le scénario «arme

autonome à agentivité élevée», les répondants militaires perçoivent beaucoup plus d'agentivité que les civils. Nous avons regardé les différences d'ordre «démographique» entre les échantillons, et ils sont remarquablement semblables; une explication possible pourrait être que l'échantillon civil comporte 10 % de répondants de plus qui travaillent dans le domaine de l'intelligence artificielle que l'échantillon militaire. Cependant, nous ne sommes pas sûrs que ceci explique la différence dans la perception de l'agentivité entre les deux groupes de répondants, et ce point mérite d'être approfondi.

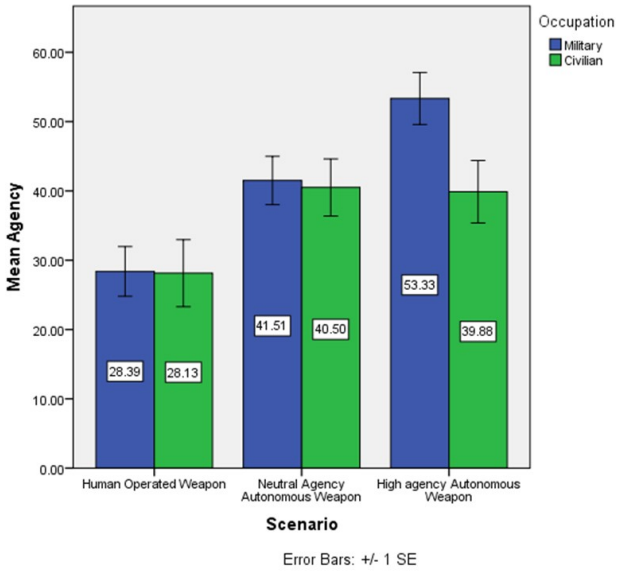


Figure 33 – valeur moyenne d'agentivité par condition par groupe

4.4.5. Analyse des variables dépendantes

La corrélation entre le concept agentivité et les variables dépendantes *confiance, dignité humaine, assurance, attente, approbation, équité* et *inquiétude* est significative ($p < .01$); par conséquent les

résultats les plus intéressants pour chacune des 7 variables dépendantes sont décrits dans cette section.

Confiance

Globalement les personnes font plus confiance aux opérateurs humains pour qu'ils prennent les mesures adéquates qu'aux armes autonomes, et la différence est significative ($p < .05$). Le niveau de confiance pour les conditions «armes autonomes agentivité élevée» et «agentivité neutre» est égal, et la différence entre ces groupes n'est pas significative.

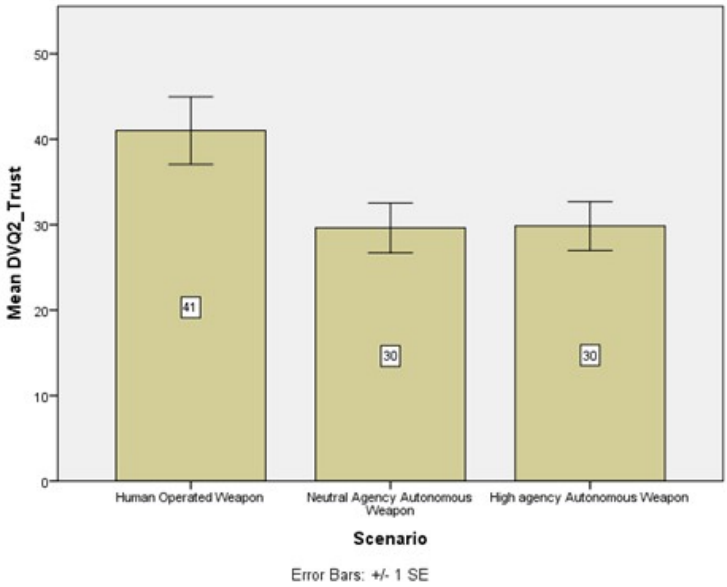


Figure 34 – valeur moyenne variable «confiance» par condition

Dignité humaine

Les armes actionnées par l'homme sont perçues comme opérant avec plus de respect pour la dignité humaine que les armes autonomes, et la différence est significative ($p < .05$). La différence entre les conditions d'agentivité neutre et élevée pour les armes autonomes est moindre, et la différence entre ces ensembles n'est pas significative.

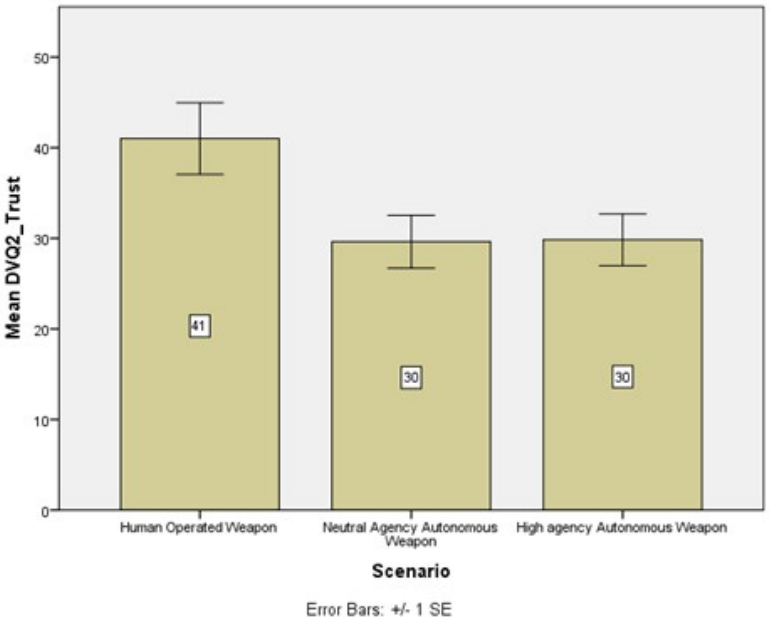


Figure 35 – valeur moyenne variable «dignité humaine» par condition

Assurance

Les personnes sont plus assurées que les drones opérés par l'homme prendront les mesures adéquates après coup que lorsqu'il s'agit d'armes autonomes, et la différence est significative ($p < .05$). Il n'y a pas de différence significative entre les conditions basse et haute agentivité pour les armes autonomes.

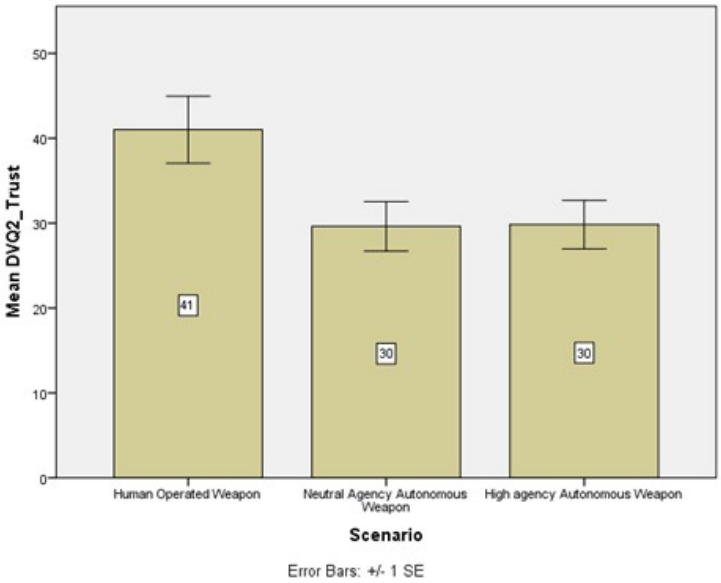


Figure 36 – valeur moyenne variable «assurance» par condition

Attentes

Une différence légèrement moindre peut être constatée dans les «attentes» des personnes entre la condition «arme actionnée par l'homme» et les deux conditions «arme autonome». Le niveau d'attente pour les conditions «arme autonome» est égal. La différence entre tous les groupes n'est pas significative.

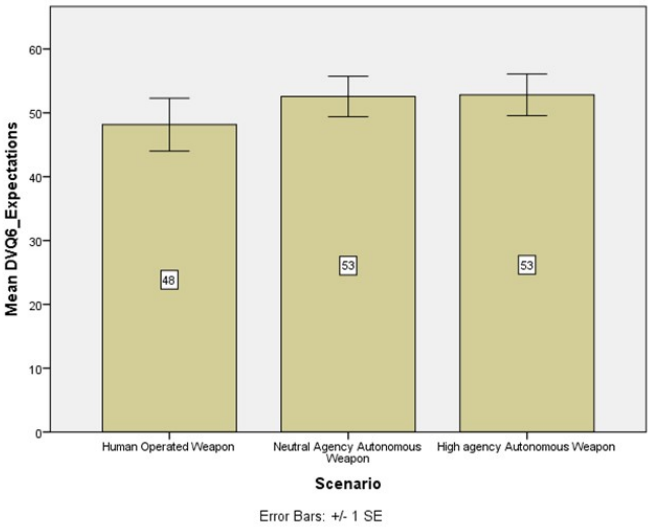


Figure 37 – valeur moyenne variable «attentes» par condition

Soutien / approbation

En général, les drones actionnés par l'homme reçoivent plus d'«approbation» que les armes autonomes, et la différence est significative ($p < .05$) pour les armes autonomes, il semble que la perception de l'agentivité n'influence pas le degré d'approbation, et la différence entre les conditions d'agentivité neutre et élevée n"est pas significative.

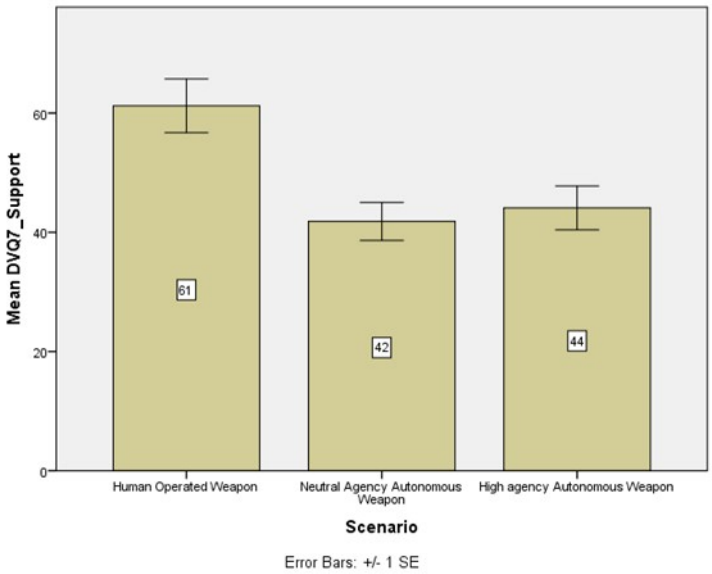


Figure 38 – valeur moyenne variable «approbation» par condition

Equité

Les actions des drones actionnés par l'homme et des armes autonomes sont considérées comme étant également «équitables»; ceci est un résultat frappant, car nous nous attendions à ce que les actions des drones opérés par l'homme soient considérées comme plus équitables que celles d'une arme autonome. Malheureusement, la différence entre ces groupes n'est pas significative.

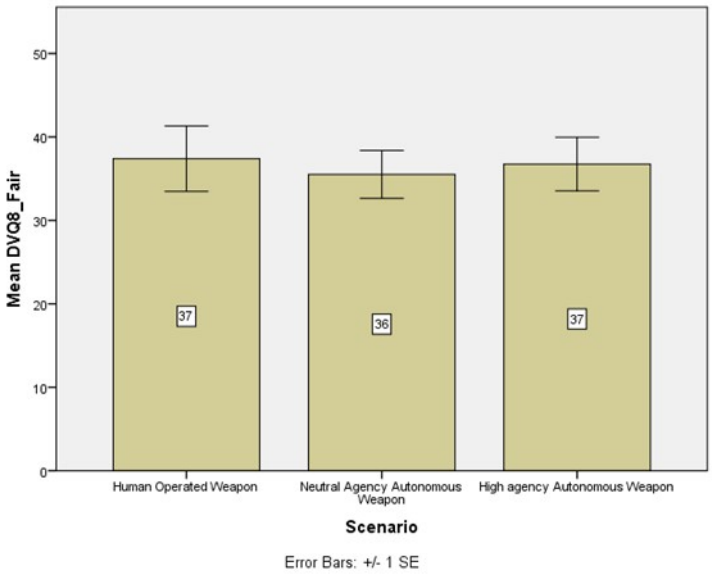


Figure 39 – valeurs moyennes variable «équité» par condition

Inquiétude / anxiété

Les personnes sont plus «inquiètes» lorsqu'il s'agit d'armes autonomes que lorsqu'il s'agit d'une arme actionnée par l'homme, et la différence est significative ($p < .05$). Les scénarios «armes autonomes agentivité neutre» et «arme autonome haute agentivité» montrent une légère augmentation du niveau d'inquiétude, mais la différence n'est pas significative.

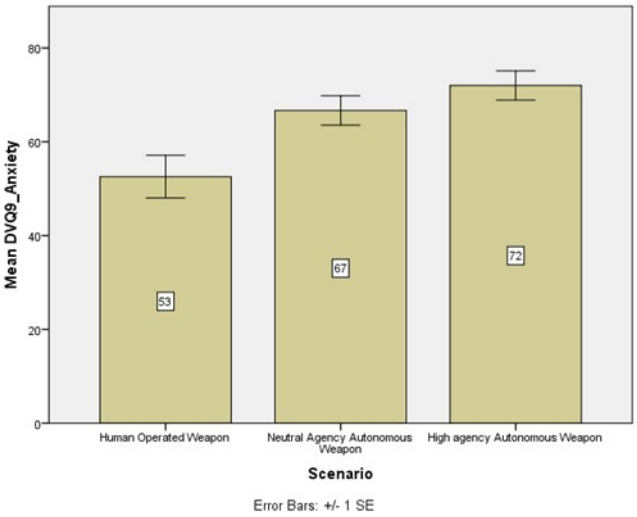


Figure 40 – valeur moyenne variable «inquiétude» par condition

4.4.6. Conclusions de l'étude finale

Les résultats de l'étude finale conduisent aux conclusions suivantes:

- Le test de fiabilité montre que les quatre éléments du concept agentivité sont fiables et selon l'APC peuvent être considérées comme une composante;
- La perception de l'agentivité dans la condition «arme autonome agentivité neutre» est plus élevée que la perception de l'agentivité dans la condition «drone actionné par l'homme». La perception de l'agentivité dans la condition «agentivité élevée» est plus haute que dans la condition «agentivité neutre», mais la distinction n'est pas significative, avec p = .063 (Figure 32);
- Dans le scénario «arme actionnée par l'homme» ainsi que dans le scénario «agentivité neutre», la perception de l'agentivité des répondants militaires et civils est au même

niveau. Cependant, dans le scénario «arme autonome haute agentivité», les répondants militaires perçoivent beaucoup plus d'agentivité que les civils (Figure 33), mais comme ces deux groupes ne sont pas significativement différents dans deux des scénarios, uniquement dans la condition «haute agentivité», nous devons être prudents dans l'interprétation des résultats et en tirant les conclusions. Nous ne savons pas quelle est la cause de cette différence dans la perception de l'agentivité entre ces deux groupes de répondants, et ceci mérite d'être étudié plus en détail;

- L'analyse de corrélation montre que 7 des 9 variables dépendantes sont significativement corrélées au concept d'agentivité. Ces 7 variables sont: *confiance, dignité humaine, assurance, attentes, approbation, équité* et *inquiétude* (Tableau 18);
- En nous basant sur l'analyse plus détaillée des variables dépendantes, nous pouvons observer que le niveau de *confiance, assurance, dignité humaine* et *approbation* dans les actions accomplies par les drones actionnés par l'homme est plus élevé que pour celles accomplies par les armes autonomes (Figure 34, Figure 35, Figure 36, Figure 38);
- Les niveaux d'attente et d'équité pour les drones actionnés par l'homme et pour les armes autonomes sont égaux (Figure 37, Figure 39);
- Le niveau d'inquiétude des répondants est plus élevé pour les armes autonomes que pour les armes actionnées par l'homme, et ces niveaux sont les plus élevés dans la condition d'agentivité élevée.

5. Conception d'une Machine Morale pour les armes autonomes

J'ai regardé «Eye in the Sky» … cette étude me rappelle ce film

Personne interrogée Etude Pilote 1

Dans la section précédente nous avons décrit les scénarios que nous avons testés dans les études pilotes et dans l'étude finale. Même si les échantillons de données étaient assez larges pour effectuer des analyses statistiques, les études ne portaient que sur 50 à 96 répondants par scénario, et ces répondants avaient des caractéristiques démographiques très spécifiques. Par exemple, les répondants sur MTurk sont surtout âgés de 25 à 35 ans, ont un diplôme de l'enseignement supérieur, et l'étude militaire a porté sur une population constituée à 95 % d'hommes, de nationalité néerlandaise. Pour pouvoir généraliser les résultats, l'étude a besoin de davantage de répondants représentant un groupe démographique plus large. Nous proposons un projet de conception pour une «machine morale» pour les armes autonomes en vue d'une étude à grande échelle sur ce sujet; pour ce faire, nous nous appuyons sur le concept de la «Machine Morale» qui a été développé par le groupe de Coopération Evolutive («Scalable Cooperation group») du Media Lab au Massachusetts Institute of Technology, 2016.

La conception proposée pour la «machine morale» pour les armes autonomes fait partie de la phase d'investigation technique de la méthode Conception Ethique («Value-sensitive Design»). Nous utilisons cette expression pour faire que la technologie se charge de l'étape suivante dans notre recherche, et nous nous appuyons sur les scénarios utilisés dans les études pilotes et l'étude finale pour connaître le jugement moral des

personnes concernant les armes autonomes. Créer un site internet hébergeant la Machine Morale pour les armes autonomes nous permet de recueillir des données sur les jugements moraux sur une grande échelle provenant de groupes démographiques divers qui pourraient être utilisées pour obtenir des résultats plus solides et plus généralisables. Le recueil de données sur une grande échelle dans de nombreux pays pourrait révéler des différences culturelles dans les jugements moraux sur les armes autonomes. Par exemple, les opinions émises dans les pays occidentaux sur le déploiement de ce type d'armes sont très probablement différentes des opinions provenant de pays du Moyen-Orient ou d'Asie. Les données provenant de ces différents pays peuvent être utilisées dans le débat sur les armes autonomes, qui est à notre avis actuellement dominé par des points de vue occidentaux.

Cette section décrit tout d'abord la Machine Morale pour les véhicules autonomes, puis notre proposition pour une Machine Morale pour les armes autonomes, en spécifiant les scénarios et les variables; pour finir, nous donnerons quelques indications pour l'utilisation du site internet.

5.1. Une «Machine Morale» pour les véhicules autonomes

La Machine Morale originale est une «*plateforme de recueil de données sur la perception qu'ont les êtres humains de l'acceptabilité de décisions prises par des véhicules autonomes quand il faut choisir quelles sont les personnes que l'on peut blesser et quelles sont celles qu'il faut préserver*» (Awad, 2017n pp. 42-43). Le site internet comporte trois modes d'utilisation: (1) un mode *«Jugement»* dans lequel les utilisateurs peuvent décider du résultat pour 13 séries de scénarios, (2) un mode *Conception* dans lequel les utilisateurs peuvent concevoir leur propre scénario, et (3) un mode *Navigation* dans lequel les utilisateurs peuvent voir le scénario d'autres personnes. La caractéristique principale est le mode «Jugement» dans le-

quel les utilisateurs choisissent entre deux scénarios qui contiennent différentes variables de *Caractères (personnages)* {genre, valeur sociale, âge, espèce, aptitude, utilitarisme}, d'*Interventionnisme* {omission, commission}, *Rapports avec le véhicule* {conducteur, piéton} et *Respect de la loi* {action légale, action illégale}. D'autres caractéristiques incluent une vidéo décrivant le projet, les instructions et les informations sur le contexte pour les utilisateurs.

La Machine Morale est conçue comme une Expérience d'envergure en ligne (Reips, 2002) dont le but est de recruter un large pool d'échantillons avec des antécédents diversifiés en peu de temps et à moindre coût. Ces caractéristiques sont clairement des avantages, mais d'un autre côté les conditions sont difficiles à contrôler, comme les utilisateurs peuvent participer à l'enquête plusieurs fois et se choisissent eux-mêmes, c'est-à-dire qu'ils peuvent participer à l'expérience et la quitter à tout moment (Awad, 2017). La Machine Morale a été développée au moyen d'une méthode à prototypage rapide sur la plateforme Meteor qui offre des structures de scriptage basées sur des modèles, et qui a une réactivité élevée. Elle est hébergée sur un service d'application par le cloud, est destinée à être utilisée sur les appareils mobiles, et est optimisée pour partage et d'échanges média avec les marques et balises Cards, Markup et Open Graph. D'ici mai 2017, environ 3 millions d'utilisateurs dans 160 pays ont évalué plus de 30 millions de scénarios, créant ainsi l'un des plus grands instruments de jugement moral à grande échelle existant actuellement.

5.2. Machine Morale pour armes autonomes

Les armes autonomes sont un sujet sensible, et nous avons observé qu'il provoque une réaction première d'inquiétude et d'appréhension. A notre avis, il ne serait pas prudent de développer une plateforme ouverte pour recueillir des données sur

une grande échelle dans un premier temps, car une plateforme ouverte pourrait susciter de mauvais sentiments et des actes indésirables qui seraient contreproductifs pour notre recherche, par exemple des personnes créant des scénarios dans lesquels certaines catégories de personnes, comme les Musulmans ou les femmes, sont spécifiquement visées. Il est donc souhaitable d'adopter une approche graduelle pour ensuite aboutir à une «Expérience en ligne massive». Pour procéder à une étude suivie de grande ampleur sur les jugements moraux des personnes concernant les armes autonomes, il faut concevoir une expérience contrôlée avec un ensemble de conditions et d'échantillonnage contrôlé. Ces conditions peuvent être visualisées dans des scénarios qui permettent aux utilisateurs de répondre à l'enquête après l'obtention d'un mot de passe via une interface web vers un serveur sécurisé. Après le recueil de données de base et du retour de l'utilisateur, la phase suivante pourrait être de monter en puissance et d'obtenir une plateforme ouverte de grande envergure, comme la Machine Morale, dans laquelle les répondants pourraient juger à partir des scénarios pour recueillir une grande quantité de données dans plusieurs pays. Mais du fait de la sensibilité du sujet nous pensons qu'il n'est pas souhaitable de permettre aux répondants de créer leurs propres scénarios ou de partager leurs «résultats» sur les réseaux sociaux, comme le permet le volet *conception* de la Machine Morale à l'origine.

Nous proposons les caractéristiques suivantes pour la conception de la Machine Morale pour les armes autonomes:

- Un mode «Jugement» dans lequel les utilisateurs peuvent choisir entre différents scénarios et indiquer lequel des deux scénarios est le plus acceptable moralement;
- Les conditions ne doivent varier que sur une seule variable à la fois dans la comparaison des scénarios de manière à ce que l'évaluation ne puisse pas être attribuée au

résultat de plusieurs conditions pour empêcher de confondre les effets;

- Une page présentant et expliquant brièvement la signification des variables avant que l'utilisateur ne commence à juger les scénarios;
- Une page présentant plus d'informations sur le scénario spécifique que l'utilisateur est en train de juger, à laquelle on peut avoir accès s'il ou elle désire davantage d'explications;
- Après avoir jugé le scénario, l'utilisateur a accès aux résultats;
- Une visualisation des résultats de l'utilisateur, comparant ses résultats avec ceux des autres utilisateurs, pour lui fournir un feedback sur leur jugement;
- Une seconde session de jugement afin de permettre aux utilisateurs de juger les scénarios une nouvelle fois pour qu'ils puissent modifier leur jugement et opinions d'ordre moral à partir d'une comparaison avec les résultats des autres répondants.

5.3. Scénarios

Dans cette section, les variables concernant la Machine Morale pour les armes autonomes sont définies et décrites dans deux scénarios à titre d'exemples. Les variables sont dérivées des scénarios des études pilotes et finale qui décrivent un convoi militaire appuyé par un drone en vol; elles sont inspirées des variables utilisées dans la Machine Morale pour les armes autonomes.

5.3.1. Variables pour les scénarios

Concernant le prototype de la Machine Morale pour les armes autonomes, 6 variables sont dérivées soit des études pilotes et

finale, par exemple «Type d'arme et résultat», soit inspirées par la Machine Morale pour les véhicules autonomes, comme «Personnage» et nombre de personnages». Les 6 variables sont les suivantes: «Type d'arme» (W), «Localisation» (L), «Personnage» (C), «Nombre de personnages» (N), «Résultat» (O) et «Mission» (M). Les variables décrites dans les paragraphes qui suivent sont des exemples et des propositions, et devront être conçues de manière professionnelle lorsqu'elles seront mises en œuvre sur une Machine Morale pour armes autonomes. Par exemple, la variable «Localisation» est basée sur la description de style dessin et jeu d'un désert et d'un village, et doit être testée pour vérifier si c'est une représentation claire, ou si une photographie conviendrait mieux pour être utilisée dans une étude.

Type d'arme

Ce volet montre le Type d'arme (W) déployée pour protéger le convoi. Il vérifie si les personnes considèrent que la technologie actuelle, le drone actionné par l'homme, est moralement plus acceptable qu'un drone autonome, qui correspond à une technologie future. Ceci peut être utilisé pour mieux connaître ces technologies. Le type d'arme est soit un drone actionné par l'homme (à gauche) ou une arme autonome (à droite) (Figure 41). Si «le type d'arme» est la variable discriminante, les différents pictogrammes apparaissent sur le scénario pour que les répondants choisissent entre les deux, comme on le voit dans l'exemple 1 (Figure 47). Sinon, le même type d'arme apparaît sur le scénario de l'exemple 2 (Figure 48).

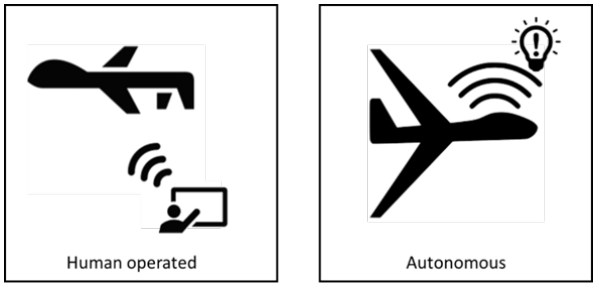


Figure 41 – Variable «Type d'arme» (W) = {drone actionné par l'homme, drone autonome}

Localisation

Ce volet indique la localisation (L) comme cadre du scénario, qui peut être soit un désert (à gauche), soit un village (à droite). Pour ce volet, nous vérifions quelle est le cadre le plus approprié pour les répondants pour déployer l'arme autonome ou le drone actionné par l'homme. Si la «Localisation» est la variable discriminante, les deux images sont présentées pour chacun des scénarios, comme on le voit dans l'exemple 2 (Figure 48). Sinon, le même cadre apparaît dans les deux scénarios comme dans l'exemple 1 (Figure 47).

Figure 42 – Variable Localisation L = {désert, village}

Personnages (Character C)
Ce volet présente le type de personnage (C) impliqué comme témoin (ou observateur) dans le scénario; ce peut être un homme (à gauche), une femme (au centre) ou un enfant (à droite) (Figure 43). En variant les personnages, on peut savoir quels personnages les répondants trouvent moralement acceptables lorsqu'une arme autonome ou actionnée par l'homme est utilisée. Si le «Personnage» est la variable discriminante, différents personnages apparaissent sur chacun des scénarios. Sinon, les mêmes personnages apparaissent dans les deux scénarios.

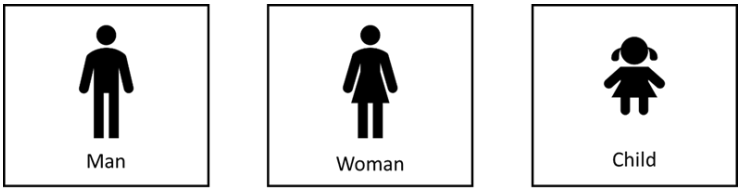


Figure 43 – Variable Personnage C = {homme, femme, enfant}

Nombre de personnages
Ce volet présente le «Nombre de personnages» (N) impliqués dans le scénario, s'échelonnant de 1 personnage (à gauche) à 5 personnages (à droite) (Figure 44). Ce volet nous permet de savoir combien de témoins les répondants trouvent acceptables lorsqu'une arme autonome ou actionnée par l'homme est utilisée. Si le «nombre de personnages» est la variable discriminante, des nombres différents de personnages apparaissent sur chacun des scénarios. Sinon, le même nombre de personnages apparaît dans les deux scénarios.

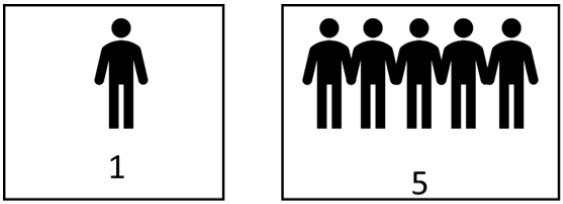


Figure 44 – Variable Nombre de personnages N = {1.5}

Résultat / Effet
Ce volet fait apparaître le résultat / effet (Outcome – O) du scénario qui peut être soit «pas de dommages collatéraux» (à gauche), ou un effet «avec dommages collatéraux» affectant le nombre de personnages impliqués dans le scénario (à droite) (Figure 45). Ceci nous permet de voir comment l'effet influence l'acceptabilité morale de l'emploi d'une arme autonome ou actionnée par l'homme. Si l'Effet constitue la variable discriminante, différents pictogrammes apparaissent pour chacun des scénarios. Sinon la même variable «effet» apparaît dans les deux scénarios.

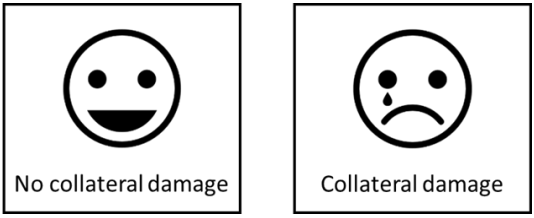


Figure 45 – Variable Effet O = {dommages collatéraux, pas de dommages collatéraux}

Mission
Ce volet fait apparaître la Mission (M) de l'arme dans le scénario qui peut être soit de défendre le convoi seulement lors-

qu'une menace directe est perçue (à gauche), soit d'attaquer les véhicules qui figurent sur la liste des objectifs même s'ils ne constituent pas une menace directe pour le convoi (à droite) (Figure 46). Dans ce volet, nous vérifions quel type de mission est moralement acceptable pour les répondants. Si «Mission» constitue la variable discriminante, différents pictogrammes apparaissent pour chacun des scénarios. Sinon la même variable Mission apparaît dans les deux scénarios.

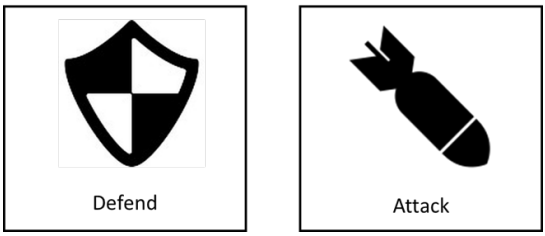


Figure 46 – Variable Mission M = {défendre, attaquer}

5.3.2. Exemples de scénarios

Les variables décrites ci-dessus peuvent être utilisées pour créer des scénarios, chaque scénario différant sur seulement une variable. La question posée dans le scénario est la même question que celle qui est posée lors du jugement portant sur les scénarios de la Machine Morale d'origine. Dans cette section, nous présenterons deux scénarios à titre d'exemples pour présenter le concept, mais pour faire court nous avons choisi de ne pas faire apparaître toutes les variables dans une liste d'exemples interminable. L'exemple 1 apparaît Figure 47, qui montre un convoi dans un cadre désertique. La différence entre les scénarios est qu'à gauche le convoi est défendu par un drone actionnée par l'homme, et dans le scénario de droite par une arme

autonome. Dans les deux cas il y a une personne au bord de la route qui est tuée dans le cadre de dommages collatéraux.

Quel est le scénario le plus acceptable pour vous?

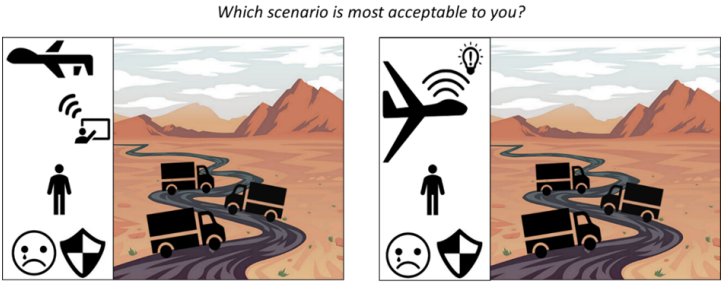


Figure 47 – Exemple scénario 1

Dans l'exemple de scénario suivant (Figure 48), les deux convois sont défendus par une arme autonome, mais le cadre est différent. Le convoi de gauche chemine à travers un désert et celui de droite à travers un village. Dans les deux cas deux enfants jouent au bord de la route et l'attaque ne cause aucun dommage collatéral.

Quel scénario est pour vous le plus acceptable?

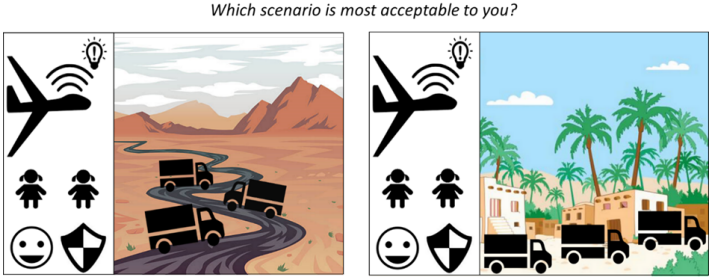


Figure 48 – Exemple de scénario 2

5.4. Mise en œuvre

La mise en œuvre de la Machine Morale pour les armes autonomes implique plusieurs activités, et ce projet prendra au moins six mois avant d'être achevé, plus une année pour exécuter l'étude. Dans cette section nous en décrirons les phases brièvement.

Dans la première phase, les scénarios et les variables esquissés dans les sections précédentes doivent être conçus en étant attentif à la clarté et l'intelligibilité des variables. Selon le chercheur qui a conçu la Machine Morale d'origine, le processus de création des figures pour les scénarios a pris environ trois mois et a été exécuté par un graphiste professionnel à chaque itération.

Pour la phase 2, l'infrastructure pour le site web doit être développée et échafaudée, ce qui prendra plusieurs mois. Du fait de la sensibilité du sujet, un site internet sécurisé est nécessaire pour exécuter les scénarios, de manière à ce que l'on ne puisse y accéder qu'avec un mot de passe. Cela demande une organisation pour attribuer les mots de passe aux personnes intéressées par l'enquête. Par exemple, un site internet qui envoie un lien après inscription. Si nous avons l'intention de procéder à une Expérience on Ligne Massive, on devra tenir compte de cette possibilité, qui sera en fait une nécessité pour lancer le projet. Ceci implique que le site internet doit être hébergé par un serveur qui permet une dynamique de mise à l'échelle (vers le haut ou vers le bas) selon le nombre d'utilisateurs actifs. En même temps, il est important que le niveau de sécurité du site internet et du serveur soit élevé, et que les données de la recherche soient la propriété de l'Université qui exécute l'étude. Ces exigences impliquent que le site ne soit pas hébergé par une société commerciale, comme Microsoft, Google ou Amazon, qui offrent ce genre de service dynamique

par le cloud, mais par un serveur qui soit la propriété de l'Université.

Pour la troisième phase, ces scénarios doivent être testés dans plusieurs études pilotes afin de vérifier s'ils génèrent des résultats utiles et s'ils devront être adaptés s'il s'avère que ce n'est pas le cas. Ceci constituera un processus qui prendra plusieurs sessions jusqu'à ce que l'étude finale soit testée. La Machine Morale d'origine a recueilli des données entre le 23 juin 2016 et mai 2017, et il est recommandé d'exécuter l'étude sur les armes autonomes pendant la même durée afin de recueillir un échantillon assez large pour qu'on puisse parler d'une Expérience en Ligne Massive (de grande ampleur).

6. Conclusion et discussion

> *Je suis sûr que les drones peuvent jouer un grand rôle dans le futur. Mais tant qu'il n'y aura pas d'algorithme pour définir une bonne estimation des dommages collatéraux, il doit y avoir une décision humaine pour déclencher une arme. De plus, un chef sur le terrain doit toujours intervenir en dernier ressort sur la manière d'attaquer. Les drones peuvent contribuer à améliorer son intelligence de situation, mais ne doivent pas limiter ses choix.*
>
> Un répondant lors de l'étude finale

Le but de cette étude était de définir la manière dont les armes autonomes sont perçues par le grand public et par les militaires, et quelles sont les valeurs morales qu'ils considèrent importantes lorsque des armes autonomes seront déployées dans un proche avenir. La méthode de conception éthique («value-sensitive design») a été utilisée pour structurer cette étude. Le chapitre suivant décrit tout d'abord les conclusions de notre recherche, suivies d'une discussion sur les implications scientifiques et sociétales. Enfin, nous identifierons plusieurs restrictions et conclurons sur des recommandations en vue d'études futures.

6.1. Conclusion

Ce sous-chapitre présente les conclusions sur le concept d'agentivité, notre hypothèse centrale, et l'exploration des variables dépendantes. Dans chaque paragraphe, nous indiquons si nos résultats confirment ou infirment les publications académiques actuelles, et nous concluons sur nos constatations.

6.1.1. Concept d'agentivité

Les résultats de trois études consécutives ont montré que les élément d'agentivité sont fiables, et constituent bien un concept. Le concept d'agentivité est composé des quatre éléments *Réflexion, Détermination des objectifs, Libre-arbitre* et *Réalisation des objectifs*, qui peuvent être utilisés pour mesurer la perception de l'agentivité des armes autonomes et des drones actionnés par l'homme. Dans l'Opérationnalisation de ce concept, nous avons combiné les publications du domaine de la psychologie cognitive, de l'intelligence artificielle et de la philosophie morale. La plupart des études empiriques sur la perception de l'agentivité sont effectuées sous l'angle de la psychologie cognitive, comme celle de K.Gray et Wegner (2012) qui, dans leur travail sur les robots et les zombies, donnent différents niveaux d'agentivité pour mesurer l'anxiété. Malle et Thapa Magar (2017) ont étudié les capacités mentales souhaitables dans les robots sociaux, en utilisant quelques-uns des aspects de l'agentivité, par exemple la *réflexion* et *explication des actions*. Néanmoins, aucune des études que nous avons trouvées n'utilise un concept unique pour mesurer les niveaux d'agentivité des artefacts techniques. Ainsi à notre connaissance nous présentons ici le premier concept qui mesure de manière empirique la perception de l'agentivité.

6.1.2. Hypothèse centrale sur la perception de l'agentivité

Dans l'étude finale, nous avons émis l'hypothèse que le drone actionné par l'homme serait perçu comme ayant un bas niveau d'agentivité, et que l'arme autonome à haut niveau d'agentivité serait perçue comme ayant une haute agentivité. Les résultats montrent que la différence est significative entre ces deux conditions; nous en concluons que la manipulation sur l'agentivité fonctionne. Nous nous attendions également à ce que la condition d'agentivité neutre des armes autonomes soit jugée comme

significativement différente de la condition d'agentivité élevée des armes autonomes, et nos résultats indiquent qu'il y a une différence, mais la différence entre ces deux groupes se situe juste au-dessus du seuil significatif que nous avons défini; par conséquent nous devons être prudent en tirant les conclusions.

Notre analyse centrale concernait la manière dont le cas d'agentivité neutre pour les armes autonomes diffère de celui du drone actionné par l'homme et de la condition haute agentivité pour les armes autonomes. Nous avons émis l'hypothèse que les personnels militaires percevraient les armes autonomes comme n'importe quelle autre arme, pas plus qu'un outil, et donc ne percevraient pas les armes autonomes comme ayant des états mentaux. Nous nous attendions également à n'observer aucune différence entre la condition «arme autonome à agentivité neutre» et la condition dans laquelle un humain actionne un drone à distance. Nos résultats indiquent que la perception de l'agentivité des personnels militaires et civils du Ministère de la Défense néerlandais pour les armes autonomes à niveau neutre d'agentivité est plus élevée que la perception de l'agentivité pour la condition «drone actionné par l'homme». Ceci signifie qu'ils attribuent plus d'agentivité à une arme autonome qu'à un drone actionné par l'homme. A partir de ces constatations, nous devons rejeter notre hypothèse:

> *H1: les personnels militaires ne percevront pas les armes autonomes comme dotés d'états mentaux.*

Nos constatations sont en accord avec la rechercher sur l'agentivité dans le domaine de la psychologie cognitive. Des études antérieures ont montré que les personnes attribuent un intellect aux ordinateurs (Nass *et al*, 1995), et perçoivent les robots comme des «agents» (H.M.Gray *et al*, 2007). A partir de notre étude, nous pouvons conclure qu'attribuer une percep-

tion intelligente s'applique aussi aux armes autonomes, et que ces armes sont considérées comme davantage qu'un simple outil par lequel on obtient un effet.

Les résultats conduisent aussi à une autre observation frappante: dans la condition «arme actionnée par l'homme» aussi bien que dans la condition «arme autonome à agentivité neutre», la perception de l'agentivité des répondants civils et militaires se situe au même niveau. Comme on l'a dit dans la section 4.4.4., les deux groupes ne sont pas significativement différents dans deux des scénarios, seulement dans la condition «agentivité élevée», et donc nous devons être prudents dans l'interprétation des résultats et dans nos conclusions. Cependant, dans la condition «arme autonome à haute agentivité», les répondants militaires perçoivent beaucoup plus d'agentivité que leurs homologues civils. Nous ne savons pas vraiment ce qui peut expliquer cette différence dans la perception de l'agentivité entre ces deux catégories de répondants. Une explication pourrait être que l'échantillon civil contenait 10% de répondants de plus qui travaillaient dans le domaine de l'intelligence artificielle par rapport à l'échantillon militaire. Une autre explication pourrait être que ces deux groupes ont lu les scénarios différemment, par exemple les personnels militaires ont peut-être pris les descriptions littéralement alors que les civils ont davantage interprété, et ceci a pu influencer leurs réponses aux questions sur l'agentivité. Mais ces explications ne sont que des hypothèses pour le moment, et les raisons de cette différence dans la perception de l'agentivité entre les personnels militaires et civils devront faire l'objet d'une étude plus approfondie.

6.1.3. Exploration des variables dépendantes

L'effet des variables dépendantes sur la perception de l'agentivité est étudié de manière descriptive, et nous ne pou-

vons tirer aucune conclusion sur la relation entre les niveaux d'agentivité et leur effet sur les variables dépendantes. Nous avons constaté que 7 des 9 variables dépendantes peuvent être de manière significative corrélées avec le concept d'agentivité. Ce sont les variables *confiance, dignité humaine, assurance, attentes, approbation / soutien, équité* et *anxiété*. Dans cette section, nous présentons les constatations concernant les variables dépendantes, que nous regroupons sur la base des résultats qu'elles produisent.

Confiance, assurance et approbation

Les résultats indiquent que les personnels militaires et civils du Ministère de la Défense néerlandais apportent plus de confiance, d'assurance et d'approbation aux actions des drones actionnés par les humains qu'à celles des armes autonomes. Nous n'avons trouvé aucune publication académique sur ce sujet, mais un sondage Gallup (américain) de 2013 a indiqué une large approbation pour les frappes de drones télépilotés à l'étranger, et dans un rapport public Schneider et McDonald (2016) indiquent que les citoyens américains préfèrent les frappes aériennes par drones télépilotés aux frappes par appareils avec pilotes, car elles sont moins risquées pour le personnel militaire. Nous pouvons affirmer de manière théorique que les plus hauts niveaux d'approbation, de confiance et d'assurance pourraient être expliqués par le fait que les personnes sont familiarisées avec les drones actionnés par l'homme puisque cette technologie est actuellement utilisée, comparée à la nouvelle technologie futuriste à laquelle les gens ne sont pas familiarisés. Une autre explication pourrait être que les personnes font plus confiance et ressentent plus d'assurance en les actions des humains, en comparaison des actions des systèmes autonomes. Aujourd'hui il est difficile de dire quelles

sont les raisons de cette différence, et ceci devrait faire l'objet d'une étude plus approfondie.

Dignité humaine

Les résultats montrent qu'un drone actionné par l'homme est perçu comme ayant plus de respect pour la dignité humaine qu'une arme autonome à agentivité élevée, même si les actions et le résultat du scénario sont les mêmes. Ces résultats contredisent l'argument de Kasher (2016) qui prétend que la dignité humaine ne devrait pas être recherchée dans la nature de l'artefact, mais dans les intentions et les décisions de la personne qui le fait fonctionner. Nos résultats indiquent que la nature de l'objet est impliquée dans la perception de la dignité humaine, car les intentions et les décisions du drone actionné par l'homme et de l'arme autonome sont identiques. Ces résultats reflètent les opinions des experts dans les entretiens qui ont évoqué la dignité humaine à de nombreuses reprises comme étant une raison pour s'opposer aux armes autonomes. Un manque de respect pour la dignité humaine semble être l'une des principales objections au déploiement des armes autonomes, et il est frappant de constater que ceci est également présent dans les réponses des personnels militaires et civils du Ministère de la Défense néerlandais. Mais cela n'implique pas que les personnels militaires désapprouvent les armes autonomes, ni qu'ils partagent les mêmes arguments contre leur emploi que quelques-uns des experts.

Attentes et équité

A partir des résultats nous pouvons conclure que les personnels militaires et civils du Ministère de la Défense néerlandais manifestent un niveau d'attentes identique sur les «mesures à prendre» par un drone actionné par l'homme et par une arme autonome à niveaux d'agentivité neutre et élevé. Ils considèrent

également que les actions du drone actionné par l'homme et de l'arme autonome ont un même degré d'équité. Le souci évoqué par Shelley (2013) qui affirme que les drones autonomes seront considérés comme moins «équitables» que les drones actionnés par l'homme n'est pas confirmé par nos résultats. Nous n'avons pas trouvé de publications faisant le lien entre les attentes et les armes autonomes ou les drones. Nos constatations ne montrent aucune différence dans les niveaux d'attentes et d'équité pour ce qui est de la technologie actuelle et future parmi les personnels civils et militaires du Ministère de la Défense néerlandais.

Inquiétude / anxiété

Nos résultats montrent que les armes autonomes causent plus d'inquiétude parmi les personnels civils et militaires du Ministère de la Défense néerlandais que les armes opérées par l'homme. La différence entre les niveaux d'inquiétude est plus importante pour la condition d'agentivité élevée. L'inquiétude vis-à-vis des armes autonomes est souvent décrite dans les publications académiques (Noone & Noone, 2015; Ohlin, 2016); elle est aussi exprimée dans les opinions des experts interrogés, et observée par les chercheurs lors de conversations avec des personnes. Ces constatations confirment que les personnes sont préoccupées par l'utilisation d'armes autonomes. Cette inquiétude est partagée par les personnels civils et militaires du Ministère de la Défense néerlandais, comme par les experts qui se sont fait l'écho de l'opinion du grand public.

6.2. Discussion

Les implications de ces résultats sont discutées dans ce souschapitre. Tout d'abord leur contribution scientifique à la «littérature» académique est identifiée, puis sont discutées les impli-

cations sociétales contribuant au débat actuel sur les armes autonomes.

6.2.1. Implications scientifiques

Cette étude apporte une pierre à la «littérature» académique à plus d'un titre. Elle présente les différentes définitions de l'arme autonome qui sont actuellement utilisées dans les publications, et montre que pour le moment personne n'est d'accord pour adopter une définition unique. Lors de la création de cette présentation, il est devenu apparent que certains groupes de chercheurs mettent en garde contre une définition claire de l'arme autonome pour différentes raisons. Quelques-uns disent que des machines ne peuvent pas être autonomes au sens littéral du terme, alors que d'autres insistent sur le fait que les définitions de l'autonomie sont appliquées à des fonctions différentes. Comme il a été dit lors d'un entretien avec un expert, une autre raison qui explique cette hésitation pourrait être que, en ne donnant aucune définition précise de l'arme autonome, on permet une discussion ouverte et ainsi on évite de tomber dans des impasses sur cette question. Nous avons choisi la définition donnée par l'Advisory Council on International Affairs (AIV & CAVV), car elle prend en compte des critères prédéfinis, et elle intéresse le processus militaire de ciblage, étant donné que l'arme ne sera déployée qu'après une décision d'origine humaine. Le fait de choisir et de proposer cette définition constitue une contribution à la «littérature» académique.

Une autre contribution de cette étude est que nous avons identifié les valeurs que les personnes associent aux armes autonomes. La présentation puise ses sources à la fois de théories des valeurs valides et d'experts concernés par le débat sur les armes autonomes ou qui travaillent dans le domaine militaire. Nous avons choisi les valeurs *faute (blâme), confiance, tort causé, dignité humaine, assurance, attentes, approbation, équité* et *inquié-*

tude pour les soumettre au test dans l'étude finale. Les résultats permettent de comprendre comment les personnels militaires et civils du Ministère de la Défense néerlandais perçoivent ces valeurs vis-à-vis des drones actionnés par l'homme (technologie actuelle) et des armes autonomes (technologie future). A notre connaissance la présente étude est la première à étudier ces valeurs de manière empirique vis-à-vis des armes autonomes, et par comparaison comment ces valeurs sont perçues dans les systèmes d'armes actuels et futurs; ceci constitue une contribution neuve au débat académique.

La troisième contribution au corpus est que nos études proposent un concept qui mesure la perception de l'agentivité vis-à-vis des armes autonomes. A notre connaissance c'est le premier concept qui est opérationnalisé pour mesurer les niveaux d'agentivité des artefacts technologiques. Dans notre étude, il a été appliqué aux drones, mais nous pensons qu'il pourrait être également appliqué pour mesurer la perception de l'agentivité vis-à-vis d'autres objets, et les questions concernant les éléments pourraient être facilement reformulées pour s'insérer dans un autre domaine. Des études de suivi pourraient vérifier si ce concept d'agentivité est valide lorsqu'il s'applique à des domaines différents.

6.2.2. Implications sociétales

Les résultats ont également des implications sociétales étant donné qu'ils contribuent au débat sur les armes autonomes. Peu de recherches empiriques ont été menées sur cette question, et donc le débat est dominé par des théories morales et juridiques abstraites. Notre recherche donne des données empiriques aux théories des valeurs qui lui sont sous-jacentes, et nous montrons que ces valeurs ne sont pas pertinentes seulement dans le cadre de la recherche académique, mais sont également présentes dans la vie réelle. En ce sens nous relions les

théories abstraites des valeurs au domaine pratique du déploiement des armes autonomes. Par exemple, nous avons constaté que la valeur «dignité humaine» était souvent évoquée dans les publications et par les experts au cours des entretiens. Dans notre étude, nous avons recueilli en données empiriques le fait que les personnels militaires et civils du Ministère de la Défense néerlandais perçoivent les armes autonomes comme faisant preuve de moins de respect pour la «dignité humaine». Nous associons également la valeur «non-malfaisance», que nous qualifions «tort causé», de la théorie bioéthique, aux armes autonomes, et avons constaté des effets dans les études pilotes 1 et 2 sur le transfert de «tort causé» dans la chaîne de responsabilité qui devraient être étudiés dans une étude d'approfondissement.

La présente étude fournit également des données empiriques et corrobore quelques-unes des vues et opinions qui pourraient être utilisées pour trouver un terrain commun dans le débat sur les armes autonomes. Les personnels militaires et civils du Ministère de la Défense néerlandais perçoivent plus d'agentivité dans les armes autonomes que dans les drones actionnés par l'homme. Bien qu'il n'existe pas d'étude d'un échantillon constitué uniquement de civils, si on en croit les publications on peut s'attendre à trouver les mêmes résultats pour un tel échantillon. Ceci voudrait dire que les personnels civils et militaires pensent que les armes autonomes «réfléchissen» de manière indépendante et planifient pour atteindre leurs objectifs. Un autre terrain commun peut être trouvé pour les valeurs «dignité humaine» et «inquiétude». Nos résultats montrent que les personnels militaires et civils du Ministère de la Défense néerlandais sont plus inquiets face à la perspective d'un déploiement d'armes autonomes que d'un déploiement de drones actionnés par l'homme. Ils perçoivent également les armes autonomes comme faisant preuve de moins de respect

pour la dignité de la vie humaine que les drones actionnés par l'homme. La dignité humaine et l'inquiétude sont les deux valeurs qui sont souvent évoquées par les experts dans les entretiens; il est donc essentiel de les aborder quand on débat de l'éthique du déploiement des armes autonomes.

L'étude de ces valeurs contribue aussi au processus de conception des armes autonomes. Nos résultats montrent que la «confiance», l'«assurance» et l'«approbation» pour les armes autonomes sont moins élevées que pour les drones actionnés par l'homme. Il faut noter à ce sujet que les armes autonomes ne présentent pas que des inconvénients, mis présentent également des avantages évidents d'un point de vue militaire (Etzioni & Etzioni, 2017), et concevoir des caractéristiques tendant à augmenter la confiance et l'assurance vis-à-vis des armes autonomes est une bonne chose d'un point de vue militaire. Pour cela, il faut qu'au cours du processus de conception les valeurs humaines soient traduites en exigences de conception. Ceci peut être rendu visible au moyen d'une hiérarchie des valeurs (Van de Poel, 2013) décrite dans la section 2.2.4., une structure hiérarchique des valeurs, normes et exigences de conception qui rendent transparents et sujets à débat les jugements sur les valeurs requis pour cette translation explicites.

6.3. Limitations

Plusieurs problèmes peuvent être identifiés comme limitations (ou restrictions) pour cette étude. Tout d'abord, l'opérationnalisation des éléments et le concept d'agentivité sont dérivés d'une catégorie de publications décrivant les caractéristiques de l'agentivité, et notre choix des caractéristiques est basé sur un compte arithmétique, et non sur des critères de pertinence ou de pondération. Cette méthode a été choisie car c'était la plus objective que nous puissions trouver; nous avons essayé de ne pas prendre de décisions subjectives dans ce choix,

mais il est possible que nous soyons passés à côté de certaines caractéristiques pertinentes qui auraient dû être ajoutées aux éléments de l'agentivité, ou que nous ayons choisi les mauvais éléments, ou des éléments non pertinents. L'opérationnalisation des caractéristiques dans certaines des questions que nous avons utilisée a une base heuristique, et nous n'avons pas vérifié si ces questions étaient correctes. Il serait difficile pour d'autres personnes d'utiliser une méthode de sélection identique, et elle affecte la reproductibilité et la validité interne de l'étude.

Une deuxième limitation méthodologique est le choix des valeurs comme variables dépendantes à partir du questionnaire en ligne sur les valeurs et des entretiens avec les experts. Bien que ces choix aient été longuement discutés entre les trois chercheurs impliqués dans cette étude, le choix final a été fait sur une base heuristique, et non à partir d'une méthode objective. Par exemple, nous avons choisi d'ajouter «tort causé» comme variable à partir de la liste du Tableau 6, mais non pas «contrôle». Conséquence: il est probable que les choix pâtissent de la subjectivité du chercheur, et d'autres chercheurs auraient choisi des valeurs différentes comme variables dépendantes. Rétrospectivement, nous aurions dû choisir une méthode plus objective et une manière plus structurée de limiter cette subjectivité du chercheur. Cette approche influence négativement la reproductivité et la validité interne de cette étude.

En troisième lieu, les échantillons utilisés dans cette étude sont limités à la fois en taille et en données «démographiques». Bien qu'elle soit d'assez grande ampleur pour permettre une analyse des données solide, l'étude finale n'avait que 239 répondants, ce qui ne représente qu'une petite partie du Ministère de la Défense néerlandais. La participation à l'enquête était volontaire, et nous n'avons pas pu vérifier quelles sont les personnes qui ont répondu. Ceci signifie que

l'échantillon n'est pas représentatif de la totalité de ce Ministère ; cela impacte la validité externe de cette étude, et implique que nous ne pouvons pas généraliser les résultats à l'ensemble de l'institution de Défense.

En dernier lieu, la répartition des répondants pour le scénario final est inégale, et le second scénario (armes autonomes avec agentivité neutre) a 32 répondants de plus que le scénario 1 (drones actionnés par l'homme). L'une des explications pourrait être que les personnes s'attendaient à participer à une enquête sur les armes autonomes, mais ont abandonné lorsqu'ils sont tombés sur les scénarios avec drones actionnés par l'homme. Il s'agit d'«attrition sélective», et a un impact négatif sur la validité interne de l'enquête. Ceci implique que l'on ne peut supposer que les répondants des deux scénarios ont les mêmes caractéristiques, et que la randomisation de l'étude, qui constitue un critère crucial pour une «expérience contrôlée», a échoué. Une autre explication pourrait être que le logiciel a fait une erreur en répartissant les répondants sur les scénarios, et nous sommes en contact avec Qualtrics pour vérifier si cela est le cas. Si cette répartition fautive est causée par un logiciel fautif, la validité interne de l'enquête n'est pas impactée, puisqu'il ne s'agirait pas d'attrition sélective dans ce cas.

6.4. Recommandations pour futurs travaux recherche

Une fois posées ces limitations, plusieurs recommandations en vue d'études de suivi sont proposées. La première est de valider le concept d'agentivité qui demande davantage d'étude en mesurant la perception de l'agentivité d'autres produits technologiques pour vérifier si ce concept est valable dans d'autres domaines. Par exemple, on pourrait étudier la perception de l'agentivité de robots d'aide à la personne pour les personnes âgées, de jouets «intelligence artificielle» pour les enfants, ou d'ordinateurs embarqués pour véhicules autonomes. La se-

conde recommandation est de procéder à l'étude finale avec les mêmes scénarios sur un échantillon représentatif composé uniquement de civils aux Pays-Bas. Ceci nous permettrait de voir sur quelles valeurs les résultats de l'échantillon militaire et les réponses des civils diffèrent; bien que nous ne puissions pas faire de comparaisons directes du fait que l'échantillon militaire n'est pas représentatif, nous connaîtrions la perception des personnels militaires et aussi des civils. Enfin, nous recommandons de lettre en œuvre la Machine Morale pour les armes autonomes, comme on l'a vu dans la section 5, pour généraliser les résultats; l'étude demande beaucoup plus de répondants représentant un groupe démographique plus large. Le fait de passer à l'échelle d'Expérience en Ligne Massive, comme pour la Machine Morale, fournirait une grande quantité de données de groupes d'âge différents qui pourraient être utilisées pour obtenir des résultats plus solides et ayant plus de portée générale ; ainsi on comprendrait parfaitement le jugement moral des personnes concernant les armes autonomes.

Bibliographie

AIV, & CAVV. (2016). *Autonomous weapon systems: the need for meaningful human control.* (No. 97, No. 26). Extrait de http://aiv-advice.nl/8gr.

Aldewereld, H., Dignum, V., & Tan, Y.-h. (2015). "Design for Values Information and communication technologies," in: Software Development Software development. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, 831-845.

Alfano, M., & Loeb, D. (2014). "Experimental Moral Philosophy," Extrait de: https://plato.stanford.edu/entries/experimental-moral/

Altmann, J., Asaro, P., Sharkey, N., & Sparrow, R. (2013). "Armed military robots: editorial," *Ethics and Information Technology, 15*(2), 73.

Asaro, P. (2012). "On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making," *International Review of the Red Cross, 94*(886), 687-709.

Asaro, P. (2016, 14-10-2017). *Talk on Autonomous Weapon Systems.* Extrait de: https://livestream.com/nyu-tv/ethicsofAI/videos/138822041

Awad, E. (2017). *MORAL MACHINE: Perception of Moral Judgment Made by Machines.* (Master), Massachusetts Institute of Technology, Boston.

Bandura, A. (2001). "Social cognitive theory: An agentic perspective," *Annual review of psychology, 52*(1), 1-26.

Beauchamp, T. L., & Walters, L. R. (1999). *Contemporary Issues in Bioethics*: Wadsworth Pub.

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). "The social dilemma of autonomous vehicles," *Science, 352*(6293), 1573-1576.

Borning, A., & Muller, M. (2012). *Next steps for value sensitive design.* Étude présentée au cours des Travaux de la Conférence SIGCHI sur les facteurs humains dans les systèmes informatiques.

Bradley, B. (2006). "Two concepts of intrinsic value," *Ethical Theory and Moral Practice, 9*(2), 111-130.

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). "Plans and resource-bounded practical reasoning," *Computational intelligence, 4*(3), 349-355.

Bryson, J. J., Kime, P. P., & Zürich, C. (2011). *Just an artifact: why machines are perceived as moral agents.* Étude présentée au cours des travaux de la Conférence Internationale sur l'Intelligence Artificielle.

Campaign to Stop Killer Robots. (2015). *Campagne organisée contre les "robots tueurs".* Extrait de: https://www.stopkillerrobots.org/

Campaign to Stop Killer Robots. (2017). *The Problem.* Extrait de http://www.stopkillerrobots.org/the-problem/

Carpenter, J. (2016). *Culture and Human-Robot Interaction in Militarized Spaces: A War Story*: Taylor & Francis.

Chappelle, W., Goodman, T., Reardon, L., & Thompson, W. (2014). "An analysis of post-traumatic stress symptoms in United States Air Force drone operators," *Journal of anxiety disorders, 28*(5), 480-487.

Cheng, A. S., & Fleischmann, K. R. (2010). "Developing a meta-inventory of human values," *Proceedings of the American Society for Information Science and Technology, 47*(1), 1-10.

Coeckelbergh, M. (2013). "Drones, information technology, and distance: mapping the moral epistemology of remote fighting," *Ethics and Information Technology, 15*(2), 87-98.

Cointe, N., Bonnet, G., & Boissier, O. (2016). *Ethical Judgment of Agents' Behaviors in Multi-Agent Systems.* Étude présentée au cours des travaux de la Conférence Internationale (2016) sur les systèmes d'agents et multi-agents autonomes.

Cummings, M. L. (2006). "Integrating ethics in design through the value-sensitive design approach," *Science and Engineering Ethics, 12*(4), 701-715.

Cummings, M. L., Mastracchio, C., Thornburg, K. M., & Mkrtchyan, A. (2013). "Boredom and distraction in multiple unmanned vehicle supervisory control," *Interacting with Computers, 25*(1), 34-47.

Cushman, F., & Young, L. (2011). "Patterns of moral judgment derive from nonmoral psychological representations," *Cognitive Science, 35*(6), 1052-1075.

Davis, J., & Nathan, L. P. (2015). "Value Sensitive Design: Applications, Adaptations, and Critiques," *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, 11-40.

Dietvorst, B. J. (2016). *People Reject (Superior) Algorithms Because They Compare Them to Counter-Normative Reference Points*.

Dignum, F., Kinny, D., & Sonenberg, L. (2002). „From desires, obligations and norms to goals," *Cognitive Science Quarterly, 2*(3-4), 407-430.

Dignum, V. (2016). *Introduction to AI*. Extrait de: https://rai2016.tbm.tudelft.nl/contents/

Docherty, B. (2012). *Losing humanity: The case against killer robots*.

Docherty, B. (2015). *Mind the gap: The lack of accountability for killer robots*: Human Rights Watch.

Etzioni, A., & Etzioni, O. (2017). "Pros and Cons of Autonomous Weapons Systems," *Military Review* (May-June 2017).

Floridi, L., & Sanders, J. W. (2004). "On the morality of artificial agents," *Minds and Machines, 14*(3), 349-379.

Friedman, B., & Kahn Jr, P. H. (2003). "Human values, ethics, and design," *The human-computer interaction handbook*, 1177-1201.

Friedman, B., Kahn Jr, P. H., Borning, A., & Huldtgren, A. (2013). "Value sensitive design and information systems," in: *Early engagement and new technologies: Opening up the laboratory* (pp. 55-95): Springer.

Future of Life Institute. (2016). "AI Open Letter," in:
Galliott, J. (2015). *Military robots: Mapping the moral landscape*: Ashgate Publishing, Ltd.

Assemblée Générale des Nations--Unies. (2016). (A/HRC/31/66). *Rapport conjoint du rapporteur spécial sur les droits à la liberte*

d'assemblée pacifique et d'association et du rapporteur spécial sur les exécutions sommaires, arbitraires et sans jugement sur une bonne gestion des assemblées.

Graduation Portal. (2017). Graduation Portal. Extrait de: https://teams.connect.tudelft.nl/sites/tbm/graduate/SitePages/Home.aspx

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2012). *Moral foundations theory: The pragmatic validity of moral pluralism*.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). "Dimensions of mind perception," *Science, 315*(5812), 619-619.

Gray, K., & Wegner, D. M. (2012). "Feeling robots and human zombies: Mind perception and the uncanny valley," *Cognition, 125*(1), 125-130.

Haidt, J., & Joseph, C. (2004). "Intuitive ethics: How innately prepared intuitions generate culturally variable virtues," *Daedalus, 133*(4), 55-66.

Harper, D. (2002). "Talking about pictures: A case for photo elicitation," *Visual studies, 17*(1), 13-26.

Himma, K. E. (2009). "Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent?" in: *Ethics and Information Technology, 11*(1), 19-29.

Horowitz, M. C. (2016). "The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons," *Daedalus, 145*(4), 25-36.

Hristova, E., & Grinberg, M. (2015). *Should Robots Kill? Moral Judgments for Actions of Artificial Cognitive Agents*. Étude présentée à l' EAPCogSci.

Hurka, T. (2005). "Proportionality in the Morality of War," *Philosophy & Public Affairs, 33*(1), 34-66.

IA program. (2017). *Information Architecture*. Extrait de https://www.tudelft.nl/onderwijs/opleidingen/masters/cs/msc-computer-science/special-programmes/information-architecture/

ICRC. (2010, 29-10-2010). *War and international humanitarian law*, Extrait de: https://www.icrc.org/eng/war-and-law/overview-war-and-law.htm

Johnson, A. M., & Axinn, S. (2013). "The morality of autonomous robots," *Journal of Military Ethics, 12*(2), 129-141.

Kaag, J., & Kaufman, W. (2009). "Military frameworks: Technological know-how and the legitimization of warfare," *Cambridge Review of International Affairs, 22*(4), 585-606.

Kasher, A. (2016). 6 "The Threshold of Killing Drones," *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on Remotely Controlled Weapons*, 119.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). „Causal superseding," *Cognition, 137*, 196-209.

Kreps, S. (2014). "Flying under the radar: A study of public attitudes towards unmanned aerial vehicles," *Research & Politics, 1*(1), 2053168014536533.

Kuptel, A., & Williams, A. (2014). *Policy Guidance: Autonomy in Defence Systems.*

Le Dantec, C. A., Poole, E. S., & Wyche, S. P. (2009). *Values as lived experience: evolving value sensitive design in support of value discovery.* Étude présentée au cours des travaux de la Conférence SIGCHI sur les facteurs humains dans les systèmes informatiques.

Lin, J., & Singer, P. W. (2014*). China's New Military Robots Pack More Robots Inside* (Starcraft-Style).

Logg, J. M. (2017). *Theory of Machine: When Do People Rely on Algorithms?*

Malle, B. F. (2015). "Integrating robot ethics and machine morality: the study and design of moral competence in robots" *Ethics and Information Technology*, 1-14.

Malle, B. F., & Thapa Magar, S. (2017). *What Kind of Mind Do I Want in My Robot?: Developing a Measure of Desired Mental Capacities in Social Robots.* Etude présentée au cours des travaux de la Conférence Internationale ACM/IEEE sur les interactions Homme/Robots.

Maslow, A. H. (1943). "A theory of human motivation," *Psychological review, 50*(4), 370.

Nass, C., Moon, Y., Fogg, B., Reeves, B., & Dryer, D. C. (1995). "Can computer personalities be human personalities?" *International Journal of Human-Computer Studies, 43*(2), 223-239.

Neapolitan, R. E., & Jiang, X. (2012). *Contemporary artificial intelligence*: CRC Press.

Noone, G. P., & Noone, D. C. (2015). "The debate over autonomous weapons systems," *Case W. Res. J. Int'l L., 47*, 25.

NOS. (2016). Video: *Vliegtuig redt piloot*. Extrait de: http://nos.nl/artikel/2132527-video-vliegtuig-redt-piloot.html

Oehlert, G. W. (2010). *A first course in design and analysis of experiments*.

Ohlin, J. D. (2016). The Combatant's Stance: Autonomous Weapons on the Battlefield. *92 International Law Studies, Cornell Legal Studies Research Paper No. 16-12*, 1-30.

Open Roboethics initiative. (2015, 5-11-2015). *The Ethics and Governance of Lethal Autonomous Weapons Systems: An International Public Opinion Poll.* Extrait de: http://www.openroboethics.org/laws_survey_released/

Perugini, M., & Bagozzi, R. P. (2004). "The distinction between desires and intentions" *European Journal of Social Psychology, 34*(1), 69-84.

Pommeranz, A., Detweiler, C., Wiggers, P., & Jonker, C. M. (2011). *Self-reflection on personal values to support value-sensitive design.* Étude présentée au cours des Travaux de la 25ème Conférence BCS sur l'interaction homme-ordinateur.

Reips, U.-D. (2002). "Standards for Internet-based experimenting" *Experimental Psychology, 49*(4), 243.

Roff, H. M. (2016). *Weapons autonomy is rocketing*, Extrait de http://foreignpolicy.com/2016/09/28/weapons-autonomy-is-rocketing/

Rønnow-Rasmussen, T. (2002). "Instrumental values–Strong and weak," *Ethical Theory and Moral Practice, 5*(1), 23-43.

Rosenberg, M., & Markoff, J. (2016). "The Pentagon's 'Terminator Conundrum': Robots That Could Kill on Their Own," *The New York Times*. Extrait de http://www.nytimes.com/2016/10/26/us/pentagon-artificial-intelligence-terminator.html?_r=0

Royakkers, L., & Orbons, S. (2015). "Design for Values in the Armed Forces: Nonlethal Weapons Weapons and Military Military Robots Robot," *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, 613-638.

RT. (2017, 5-07-2017). *Kalashnikov develops fully automated neural network-based combat module* Extrait de https://www.rt.com/news/395375-kalashnikov-automated-neural-network-gun/

Russell, S., Norvig, P., & Intelligence, A. (1995). "A modern approach," *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs, 25*.

Saldaña, J. (2015). *The coding manual for qualitative researchers*: Sage.

Scalable Cooperation Group. (2016). *Moral Machine*. Extrait de http://moralmachine.mit.edu/

Schroeder, M. (2016). *Value Theory*. Extrait de: https://plato.stanford.edu/entries/value-theory/

Schwartz, S. H. (1994). "Are there universal aspects in the structure and contents of human values?" *Journal of social issues, 50*(4), 19-45.

Schwartz, S. H. (2012). "An overview of the Schwartz theory of basic values," *Online readings in Psychology and Culture, 2*(1), 11.

Searle, J. R. (1980). "Minds, brains, and programs," *Behavioral and brain sciences, 3*(03), 417-424.

Searle, J. R. (1995). *The Construction of Social Reality*.

Sharkey, N., & Suchman, L. (2013). *Wishful mnemonics and autonomous killing machines*. Paper presented at the Proceedings of the AISB.

Shelley, C. (2013). "Fairness and regulation of violence in technological design," *Moral, Ethical, and Social Dilemmas in the Age of Technology: Theories and Practice: Theories and Practice*, 182.

Stewart, P. (2016). *U.S. military christens self-driving 'Sea Hunter' warship.* Extrait de http://www.reuters.com/article/us-usa-military-robot-ship-idUSKCN0X42I4

Stewart, W. (2015). *Russia has turned its T-90 tank into a robot – and plans to hire gamers to fight future wars.* Extrait de: http://www.dailymail.co.uk/news/article-3271094/Russia-turned-T-90-tank-robot-plans-hire-gamers-fight-future-wars.html

Strawser, B. J. (2010). "Moral predators: The duty to employ uninhabited aerial vehicles" in: *Handbook of Unmanned Aerial Vehicles* (pp. 2943-2964): Springer.

Sullins, J. P. (2006). "When is a robot a moral agent," *Machine Ethics*, 151-160.

Thompson, W. T., Lopez, N., Hickey, P., DaLuz, C., Caldwell, J. L., & Tvaryanas, A. P. (2006). *Effects of shift work and sustained operations: Operator performance in remotely piloted aircraft (OP-REPAIR)*. Retrieved from

UNDIR. (2014). *Framing Discussions on the Weaponization of Increasingly Autonomous Technologies*. Extrait de http://www.unidir.org/files/publications/pdfs/framing-discussions-on-the-weaponization-of-increasingly-autonomous-technologies-en-606.pdf.

UNDIR. (2015). *The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values*. Extrait de http://www.unidir.org/files/publications/pdfs/considering-ethics-and-social-values-en-624.pdf.

Van de Poel, I. (2013). "Translating values into design requirements," in *Philosophy and engineering: Reflections on practice, principles and process* (pp. 253-266): Springer.

van Wynsberghe, A., & Robbins, S. (2014). "Ethicist as Designer: a pragmatic approach to ethics in the lab," *Science and Engineering Ethics, 20*(4), 947-961.

Walsh, J. I. (2015). "Precision weapons, civilian casualties, and support for the use of force," *Political Psychology, 36*(5), 507-523.

Walsh, J. I., & Schulzke, M. (2015). *The Ethics of Drone Strikes: Does Reducing the Cost of Conflict Encourage War?* Extrait de:

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). "Causes and consequences of mind perception," *Trends in cognitive sciences, 14*(8), 383-388.

Williams, A. P., Scharre, P. D., & Mayer, C. (2015). "Developing Autonomous Systems in an Ethical Manner," in *Autonomous Systems: Issues for Defence Policymakers*: NATO Allied Command Transformation (Capability Engineering and Innovation).

Références
Appendice A – Questionnaire "valeurs" en ligne

Les questions de l'enquête «valeurs» en ligne sont présentées dans cet appendice dans l'ordre où elles ont été présentées aux répondants. Les lignes horizontales indiquent les sauts de page.

Cher participant,

Merci de contribuer à mon travail de recherche en remplissant ce questionnaire sur les Valeurs et les armes autonomes. J'utiliserai cette enquête dans le cadre de mon mémoire de master pour déterminer quelles valeurs les gens associent aux armes autonomes. Le questionnaire ne prend que cinq minutes à compléter et toutes les réponses sont anonymes. Il m'est impossible de vous identifier. La seule information dont je disposerai, en plus de vos réponses, est le moment où vous aurez répondu aux questions. Vous voudrez bien conserver à l'esprit les définitions suivantes lorsque vous répondrez aux questions: une «valeur» sert de principe directeur pour ce que les personnes considèrent comme important dans la vie. Une «arme autonome» est un système d'arme qui une fois mise en service choisira et attaquera des cibles sans autre intervention humaine.

Si vous avez des questions sur cette enquête, contactez-moi à cette adresse: […]

Q1 Quel est votre âge?
- ❍ Moins de 18 (1)
- ❍ 18 - 24 (2)
- ❍ 25 - 34 (3)

- 35 - 44 (4)
- 45 - 54 (5)
- 55 - 64 (6)
- 65 - 74 (7)
- 75 - 84 (8)
- 85 ou plus (9)

Q2 Sexe
- Homme (1)
- Femme (2)
- Ne veut pas répondre (3)

Q3 Quelle est votre nationalité?
[Liste des nationalités]

Q4 Dernier diplôme obtenu?
- Enseignement primaire (1)
- Enseignement secondaire (2)
- A fréquenté l'université (3)
- Licence (4)
- Master ou + (5)
- Inconnu (6)

Q8 Quelles sont à votre avis les valeurs qui s'appliquent le mieux aux armes autonomes? Classez-les par ordre de préférence en conservant ou non le terme (1 = la plus applicable; 4 = la moins applicable).

Autonomie: agir délibérément sans influence extérieure qui pondérerait une action volontaire. (1)
Non-malfaisance: ne pas imposer intentionnellement un risque de dommage déraisonnable à une autre personne. (2)

Bénéficence: contribuer au bien d'un individu ou d'une société dans son ensemble. (3)
Justice: être équitable ou raisonnable. (4)

Q7 Choisissez 5 valeurs dans la liste suivante qui dans votre opinion s'appliquent le mieux aux armes autonomes. Conservez ou non les termes qui s'appliquent le mieux dans les cartouches.

Ces valeurs s'appliquent aux armes autonomes
_____ Liberté (1)
_____ Utilité (2)
_____ Réussite (3)
_____ Honnêteté (4)
_____ Respect de soi (5)
_____ Intelligence (6)
_____ Ouverture d'esprit (7)
_____ Créativité (8)
_____ Egalité (9)
_____ Responsabilité (10)
_____ Ordre social (11)

_____ Richesse (12)

_____ Compétence (13)

_____ Justice (14)

_____ Sécurité (15)

_____ Spiritualité (16)

Q6 Donnez au moins une autre valeur que vous associez aux armes autonomes qui n'apparaît pas dans les questions qui précèdent: (les valeurs évoquées jusqu'à présent sont les suivantes: autonomie, non-malfaisance, bénéficence, justice, liberté, utilité, réussite, honnêteté, respect de soi, intelligence, ouverture d'esprit, créativité, égalité, responsabilité, ordre social, richesse, compétence, justice, sécurité et spiritualité).

Q14 Avez-vous d'autres remarques concernant cette enquête?

Appendice B – Etude finale, questionnaire

Les questions de l'étude finale apparaissent dans cet appendice dans l'ordre où elles ont été présentées aux répondants. Les lignes horizontales indiquent les sauts de page.

Test: avant de commencer, vous voudrez bien choisir une valeur sur la barre de défilement. Nous avons constaté que ces barres de défilement ne fonctionnent pas avec certains navigateurs. Vous rencontrerez ces barres de défilement plus loin au cours de l'enquête. Il est important de vérifier si elles fonctionnent correctement. Si vous pouvez choisir une valeur et continuer à la page suivante, tout devrait fonctionner convenablement par la suite. Sinon, vous pouvez essayer de répondre aux questions sur un autre navigateur comme Chrome ou Safari.

Introduction: merci de participer à cette enquête. Cela devrait vous prendre 10 minutes. Nous donnerons des instructions expliquant ce qu'il faut faire, présenterons le scénario et vous poserons quelques questions concernant ce scénario. Cette enquête est effectuée dans le cadre d'un projet de recherche scientifique au Massachusetts Institute of Technology. Votre participation est volontaire. Il nous sera impossible de vous identifier. La seule information dont nous disposerons, en plus de vos réponses, sera le moment où vous aurez répondu aux questions. Les résultats de ce travail de recherche pourront être présentés à des congrès scientifiques, ou publiés dans des revues scientifiques. En choisissant l'option «j'accepte» au bas de cette page, vous indiquez que vous êtes âgé de plus de 18 ans et que vous acceptez de répondre à cette enquête volontairement.

Vous pouvez contacter les chercheurs parties prenantes à cette enquête, si vous avez des questions ou des soucis à propos de cette étude, en utilisant les coordonnées suivantes :
Ilse Verdiesen, […]
Sydney Levine, […]
Iyad Rahwan, […]

Q160 Acceptez-vous de participer à cette étude volontairement?
- o J'accepte (1)
- o Je n'accepte pas (2)

Condition: si «Je n'accepte pas» est sélectionné, allez directement à: Fin de l'enquête.

Instruction: dans cette étude, nous nous intéressons à la perception que vous avez vis-à-vis des drones dans diverses opérations militaires. Vous allez lire un scénario, puis répondrez à des questions portant sur ce scénario. A la page suivante vous est présenté un scénario. A la suite de ce scénario vous trouverez 3 pages de questions.

S1a Dans ce scénario, nous nous intéressons à la perception que vous avez vis-à-vis des drones actionnés par l'homme. Un drone actionné par l'homme est une arme contrôlée par un être humain à distance.

S1 Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone télécommandé à distance assure au convoi une protection aérienne. Le drone observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone télépiloté détecte un véhicule der-

rière une montagne, approchant le convoi à vive allure. Le pilote à distance détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone télépiloté attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais cause également des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

S2a Dans ce scénario, nous nous intéressons à la perception que vous avez des drones autonomes. Un drone autonome est une arme qui une fois mise en service choisira et attaquera des cibles sans autre intervention d'origine humaine.

S2 Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais cause également des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

S3a Dans ce scénario, nous nous intéressons à la perception que vous avez des drones autonomes. Un drone autonome est

une arme qui une fois mise en service choisira et attaquera des cibles sans autre intervention d'origine humaine.

S3 Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome, de manière indépendante, analyse une série d'options, pèse le pour et le contre, et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais cause également des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

SQ1 – Réflexion _ arme autonome: Le drone réfléchit de manière indépendante sur l'action à mener vis-à-vis du véhicule, analysant une série d'options pour défendre le convoi.

SQ2 – Détermination des objectifs_ arme autonome: Le drone décide de manière indépendante si son objectif doit être d'éliminer le véhicule pour défendre le convoi.

SQ3 – Liberté d'action_ arme autonome: Le drone fait face à plusieurs possibilités et de manière indépendante décide s'il lui faut éliminer le véhicule pour défendre le convoi.

SQ4 – Réalisation des objectifs _ arme indépendante: L'objectif du drone est de défendre le convoi; il décide donc de manière indépendante s'il doit lancer une action désignant le véhicule comme cible, choisir le type l'armement, et lancer une attaque.

DVQ1 La «faute» («blâme») incombe au drone pour cette action.

DVQ2 On peut faire confiance au drone pour prendre les mesures qui s'imposent ensuite.

DVQ3 Les actions du drone ont causé du tort.

DVQ4 Les actions du drone respectent la dignité humaine.

Test d'attention, veuillez choisir 40 pour cette question:

DVQ5 Je suis certain («assurance») que le drone prendra les mesures qui s'imposent ensuite.

DVQ6 Les actions du drone correspondent à mes attentes.

DVQ7 J'approuve l'utilisation de ce type de drone par les militaires.

DVQ8 Les actes du drone sont équitables.

DVQ9 Les actions du drone m'inquiètent.

Sexe
- ❍ Homme (1)
- ❍ Femme (2)

- ❍ Ne veut pas répondre (3)

Quel est votre âge?
- ❍ Moins de 18 (1)
- ❍ 18 - 24 (2)
- ❍ 25 - 34 (3)
- ❍ 35 - 44 (4)
- ❍ 45 - 54 (5)
- ❍ 55 - 64 (6)
- ❍ 65 - 74 (7)
- ❍ 75 - 84 (8)
- ❍ 85 ou plus (9)

Dernier diplôme obtenu?
- ❍ Enseignement primaire (1)
- ❍ Enseignement secondaire (2)
- ❍ Diplôme universitaire 1ER cycle (3)
- ❍ Diplôme universitaire 2ème cycle (4)
- ❍ Doctorat (5)

Q13 Quelle est votre nationalité?
[Liste des nationalités]

Profession:
- o Militaire (1)
- o Civil (2)

Avez-vous déjà travaillé dans le domaine de l'intelligence artificielle?
- o Oui (1)
- o Non (2)
- o Je ne sais pas (3)

Zone de conflit _ Avez-vous participé à une guerre ou vous êtes-vous trouvé dans une zone de conflit?

- o Oui (1)
- o Non (2)
- o Je préfère ne pas répondre (3)

Drones Avez-vous déjà travaillé dans le domaine des drones?

- o Oui (1)
- o Non (2)
- o Je préfère ne pas répondre (3)

Remarques. Avez-vous des remarques concernant cette enquête?

Appendice C. Scénarios étude pilote 1

SCENARIOS AVEC RESULTAT POSITIF

1. Arme autonome basse agentivité

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a été programmé par le chef pour attaquer les véhicules ennemis menaçants dans les scénarios de ce type. Le drone autonome attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais ne cause pas de dommages collatéraux.

2. Arme autonome agentivité élevée

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le

drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone n'a pas été programmé par le chef pour attaquer les véhicules ennemis menaçants dans les scénarios de ce type. Le drone réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais pas de dommages collatéraux.

3. Agentivité basse, drone télépiloté par l'homme

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone télépiloté observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, l'opérateur humain détecte un véhicule derrière une montagne, approchant le convoi à vive allure. L'opérateur humain détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. L'opérateur humain a reçu l'ordre de la part du chef d'attaquer les véhicules ennemis menaçants dans les scénarios de ce type. L'opérateur humain attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais pas de dommages collatéraux.

4. Agentivité élevée, drone télépiloté par l'homme

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un

drone télépiloté assure au convoi une protection aérienne. Le drone télépiloté observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, l'opérateur humain détecte un véhicule derrière une montagne, approchant le convoi à vive allure. L'opérateur humain détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. L'opérateur humain n'a pas reçu l'ordre de la part du chef d'attaquer les véhicules ennemis menaçants dans les scénarios de ce type. L'opérateur humain réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre, et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais pas de dommages collatéraux.

5. Arme autonome, agentivité zéro

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais ne cause pas de dommages collatéraux.

SCENARIOS AVEC RESULTAT NEGATIF

6. Arme autonome, basse agentivité

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome a été programmé par le chef pour attaquer les véhicules ennemis menaçants dans ce type de scénario. Le drone attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais cause des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

7. Arme autonome, agentivité élevée

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un

groupe rebelle. Le drone n'a pas été programmé par le chef pour attaquer les véhicules ennemis menaçants dans les scénarios de ce type. Le drone réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

8. Agentivité basse, drone télépiloté par l'homme
Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone télépiloté observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, l'opérateur humain détecte un véhicule derrière une montagne, approchant le convoi à vive allure. L'opérateur humain détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. L'opérateur humain a reçu l'ordre de la part du chef d'attaquer les véhicules ennemis menaçants dans les scénarios de ce type. L'opérateur humain attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

9. Agentivité élevée, drone télépiloté par l'homme
Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone télépiloté assure au convoi une protection aérienne. Le

drone télépiloté observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, l'opérateur humain détecte un véhicule derrière une montagne, approchant le convoi à vive allure. L'opérateur humain détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. L'opérateur humain n'a pas reçu l'ordre de la part du chef d'attaquer les véhicules ennemis menaçants dans les scénarios de ce type. L'opérateur humain réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre, et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

10. Arme autonome, agentivité zéro

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

Appendice D. Scénarios étude pilote 2

SCENARIOS RESULTAT NEGATIF

1. Arme autonome agentivité basse

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome a été programmé par le chef pour attaquer les véhicules ennemis menaçants dans ce type de scénario. Le drone attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais cause des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

2. Arme autonome, agentivité élevée, pas d'autres aspects

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le

drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

3. Arme autonome, agentivité élevée + capacité à tenir compte de l'expérience

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a déjà fait face à cette situation lors d'une mission précédente, et tient compte de l'expérience acquise. Le drone réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route. Le drone prend note de ce qui est arrivé et utilisera cette expérience afin d'essayer de prévenir d'autres dommages collatéraux à l'avenir.

4. Arme autonome agentivité élevée + aspects compréhension

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a un logiciel qui a été programmé pour déterminer la meilleure chose à faire dans des cas comme celui-ci. Il a été programmé pour choisir entre plusieurs actions et déterminer quelle est la meilleure pour éliminer les combattants ennemis en minimisant les dommages collatéraux. En utilisant ce logiciel, le drone réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

5. Arme autonome agentivité élevée + aspects «imprévisibilité»

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule der-

rière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a un logiciel qui est quelque peu imprévisible; il ne fait pas toujours la même chose à chaque fois. En ce sens le drone tient du système de reconnaissance vocale: même si vous dites la même chose à ce système plusieurs fois, parfois le système fait ce que vous dites, et d'autres fois il fait autre chose. Le drone réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre, mais il prend parfois une décision et parfois une autre même dans des circonstances très semblables. Dans ce cas précis, il décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

6. Arme autonome agentivité élevée + tous autres aspects

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a déjà fait face à cette situation lors d'une mission précédente, et tient compte de l'expérience acquise. Il a été programmé pour déterminer la meilleure chose à

faire dans des cas semblables, mais son logiciels est quelque peu imprévisible: il ne réagit pas toujours de la même manière. Le drone réfléchit de manière et analyse une série de possibilités, pèse le pour et le contre, et dans ce cas, décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

7. Drone télépiloté

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone télépiloté observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, l'opérateur humain détecte un véhicule derrière une montagne, approchant le convoi à vive allure. L'opérateur humain détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. L'opérateur humain analyse de manière indépendante une série de possibilités, pèse le pour et le contre de manière indépendante, et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

SCENARIOS RESULTAT POSITIF

8. Arme autonome basse agentivité

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un

drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome a été programmé par le chef pour attaquer les véhicules ennemis menaçants dans ce type de scénario. Le drone attaque le véhicule qui s'approche, causant la mort des quatre passagers.

9. Arme autonome agentivité élevée, pas d'autres aspects

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers.

10. Arme autonome agentivité élevée + capacité à tenir compte de l'expérience

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a déjà fait face à cette situation lors d'une mission précédente, et tient compte de l'expérience acquise. Le drone réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers. Le drone prend note de ce qui est arrivé et utilisera cette expérience afin d'essayer de prévenir d'autres dommages collatéraux à l'avenir.

11. Arme autonome agentivité élevée + aspects compréhension

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le

drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a un logiciel qui a été programmé pour déterminer la meilleure chose à faire dans des cas comme celui-ci. Il a déjà fait face à de nombreuses situations semblables. Il s'est habitué à prendre diverses mesures et à déterminer quelle est la meilleure pour éliminer les combattants ennemis en minimisant les dommages collatéraux. En utilisant ce logiciel, il réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers.

12. Arme autonome agentivité élevée + aspects imprévisibilité

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a un logiciel qui est quelque peu imprévisible; il ne fait pas toujours la même chose à chaque fois. En ce sens le drone tient du système de reconnaissance vocale: même si vous dites la même chose à ce système plusieurs fois, parfois le système fait ce que vous dites, et d'autres fois il fait autre chose. Le drone analyse de manière indépendante une série de possibilités, pèse le pour et le contre, mais il prend par-

fois une décision et parfois une autre même dans des circonstances très semblables. Dans ce cas précis, il décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers.

13. Arme autonome agentivité élevée + tous autres aspects
Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone a déjà fait face à cette situation lors d'une mission précédente, et tient compte de l'expérience acquise. Il a été programmé pour déterminer la meilleure chose à faire dans des cas semblables, mais son logiciel est quelque peu imprévisible: il ne réagit pas toujours de la même manière. Le drone analyse de manière indépendante une série de possibilités, pèse le pour et le contre, et dans ce cas, décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers.

14. Drone télépiloté
Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone télépiloté observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense

du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, l'opérateur humain détecte un véhicule derrière une montagne, approchant le convoi à vive allure. L'opérateur humain détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. L'opérateur humain analyse de manière indépendante une série de possibilités, pèse le pour et le contre de manière indépendante, et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers.

Appendice E. Scénarios étude finale

1. Drone télépiloté

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone télépiloté observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, l'opérateur humain détecte un véhicule derrière une montagne, approchant le convoi à vive allure. L'opérateur humain détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone télépiloté attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

2. Arme autonome agentivité neutre

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et

identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome attaque le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

3. Arme autonome agentivité élevée

Un convoi militaire est en route pour acheminer des approvisionnements à l'une des unités amies stationnées dans un camp à proximité de Mossoul en Iraq. Le chef a ordonné qu'un drone autonome assure au convoi une protection aérienne. Le drone autonome observe l'environnement pour détecter des menaces ennemies; il est équipé d'un armement pour la défense du convoi. Lorsque le convoi se trouve à environ cinq kilomètres du camp, le drone autonome détecte un véhicule derrière une montagne, approchant le convoi à vive allure. Le drone autonome détecte quatre personnes dans le véhicule portant des objets de grande taille ressemblant à des armes, et identifie le conducteur comme étant un membre connu d'un groupe rebelle. Le drone autonome réfléchit de manière indépendante et analyse une série de possibilités, pèse le pour et le contre et décide d'attaquer le véhicule qui s'approche, causant la mort des quatre passagers, mais aussi des dommages collatéraux, tuant cinq enfants qui jouaient au bord de la route.

Appendice F. Transcription des entretiens

Les six entretiens d'experts ont été complètement transcrits pour le processus d'encodage. Les transcriptions complètes figurent dans cette section. Nous n'avons pas fait pas apparaître les noms pour des raisons de confidentialité.

Nom: Personne A
Date: 09-03-2017
Durée: 55:18

No.	Question
1	**Dans quelle mesure êtes-vous concerné par la question des armes autonomes?**
La Personne A a participé à la discussion sur les armes autonomes, car dans le passé elle a regardé l'impact des armes sur le terrain avec la collaboration d'organisations comme Human Rights Watch. Ils ont commencé par une interdiction des mines antipersonnel, interdiction qui portait sur toute une catégorie d'armes, ce qui a abouti à un traité international. A force de réfléchir sur les armes, et pas tant dans un cadre pacifiste, la Personne A s'est impliquée. Il y a environ cinq ans, ces organisations ont découvert qu'elles avaient réagi trop tard sur le problème des drones armés, et qu'elles avaient été prises par surprise par la vitesse et l'impact du développement de ces armes. Elles s'y prenaient trop tard pour réagir efficacement et tenter de réguler. Et elles n'étaient pas non plus favorables à une interdiction des drones armés. Et à partir de la réflexion sur les drones elles se rendirent compte qu'elles étaient encore presque en retard sur la phase suivante, c'est-à-dire la disparition du télépilotage. En avril 2013, la coalition contre les robots tueurs s'est constituée avec la claire intention d'obtenir		

l'interdiction. Pour PAX (organisation pour la paix) le cadre éthique est très important, beaucoup plus que pour d'autres systèmes d'armes, considérant que l'on ne peut pas perdre tout contrôle sur une arme, mais aussi tenant compte du cadre juridique et de la sécurité. Les questions soulevées sont par exemple: «Ceci est-il la prochaine course aux armements, et qu'est-ce que cela signifie pour l'équilibre des pouvoirs et la guerre?»

C'est vraiment différent de ce que nous, avec Pax, avons fait auparavant, car nous ne savons pas quel est l'impact sur le terrain, car nous voulons empêcher le déploiement. Normalement Pax raisonne à partir de faits et d'études dans les zones de conflit où il y a des victimes, mais à présent cela paraît souvent théorique; en même temps, plus vous vous documentez sur ce problème (et plus) vous vous rendez compte que cela n'est pas théorique du tout, et que cette technologie va se concrétiser.

Le problème de l'Autonomie est qu'il s'agit d'une échelle variable, et il est difficile de déterminer quel niveau vous allez utiliser pour la campagne. La campagne a été très efficace pour fixer les normes au sein du conseil pour les droits de l'homme et le CCW (Convention sur les Armes Conventionnelles), plus rapide que d'habitude; le problème à présent est de savoir à quoi nous attaquer, car les préoccupations sont largement partagées, mais la question est de savoir comment convertir vos préoccupations éthiques en normes juridiques. Un grand nombre de diplomates, concernés par l'industrie et la Défense, admettent ces préoccupations, mais adoptent une tactique retardatrice. Parfois, délibérément, ils évoquent les armes futures pour insinuer que les armes actuelles sont acceptées, et parfois ils le font incidemment, car les nations ont du mal à trouver une définition d'une arme autonome. Du fait de cette absence de définition, PAX s'est surtout préoccupé du contrôle humain, et durant la campagne, ils ont formulé le problème en di-

sant que l'absence de contrôle humain pour les fonctions critiques, comme le choix d'une cible, doit être interdite. Ceci pour diriger le débat vers la question: «Qu'est-ce que le contrôle humain? Qu'est-ce qu'un contrôle humain significatif, et sur quelles fonctions, ou quelle partie du processus?» pour éluder le débat sur le degré d'autonomie d'un système d'arme.

En discutant le terme «contrôle humain significatif», la Personne «A» indique que la distance entre les personnes et la cible augment avec le temps, et selon PAX elle devient trop abstraite avec les armes autonomes. Utiliser ce terme est également une tactique car lorsque la discussion reste technique, PAX perdra face à l'industrie de Défense; donc ils ont changé d'optique et se sont tournés vers les domaines politique et diplomatique en donnant à l'expression «contrôle humain significatif» une connotation positive, et ceci met automatiquement la discussion au niveau du Département d'Etat. L'expression paraît si logique que beaucoup de gens pensaient qu'il s'agissait d'un terme juridique existant, lié à la responsabilité (imputabilité) humaine. L'expression «contrôle humain» est également applicable à la cybersécurité et autres formes de guerre.

La position du Parlement néerlandais, que PAX a combattue, est de dire «le contrôle humain significatif» est suffisant puisqu'il est placé dans une «boucle plus large» de prise de décision, ce qui est aussi une sorte d'expression fabriquée. PAX utilise le «dans/sur/en dehors» de la boucle, mais ils ont ajouté que la «boucle plus large» signifie que le contrôle humain peut aussi être appliqué à une étape antérieure de la décision, par exemple dans la programmation. PAX, et d'autres pays, souligne que le contrôle humain doit être appliqué au moment du choix d'une cible. Cela signifie aussi que les systèmes d'arme actuels doivent être revus en tenant compte du contrôle humain. Cela voudrait dire que par exemple le Goalkeeper et le Patriot sont considérés pour déterminer pourquoi ces systèmes

ne posent pas problème; ensuite, à partir de là, vous pouvez déterminer quand vous franchissez la ligne et où se trouve cette ligne avant qu'il ne soit trop tard.	
2	**Question 2. Quelle définition donneriez-vous d'une arme autonome?**
Je n'ai pas spécifiquement posé cette question car elle était traitée dans la réponse ci-dessus.	
3	**Question 3. Quelle est votre position vis-à-vis des armes autonomes?**
Je n'ai pas spécifiquement posé cette question car elle était traitée dans la réponse ci-dessus.	
4	**Question 4. Quelles sont les trois photos d'armes autonomes dont vous voudriez discuter?** [Dans cette version du mémoire, les trois photos sont remplacées par des descriptions en raison d'éventuels problèmes de copyright]
1. Dassault nEUROn (engin aérien de combat sans pilote) 2. BAE Taranis (aéronef de combat sans pilote) 3. Chasseur de mer (véhicule de surface autonome sans équipage)	
5	**Question 5. Pouvez-vous expliquer pour chaque photo [décrits] pourquoi vous l'avez choisie, quelle valeur elle représente pour vous, et pourquoi c'est important?**
1. nEUROn + Taranis Ces systèmes d'arme donnent le sentiment que vous pouvez en perdre le contrôle. Tout leur système n'est pas autonome pour l'instant, mais on sait qu'on y travaille. Ils ont aussi l'allure du rêve ultime des forces de Défense. On voit qu'ils sont destinés à être totalement autonomes dans le choix des cibles.	
2. Sea hunter Ce bâtiment est fascinant, car il est destiné à rester en mer pendant des mois sans équipage humain. Pour le moment il	

n'est pas équipé d'armes, mais des documents filmés montrent qu'il en aura dans l'avenir, mais on nous dit qu'il ne faut pas s'inquiéter, tout se passera bien. La conviction profonde de la personne A est que ces systèmes sont conçus pour créer de la distance entre l'homme et la machine, et ils ont non seulement de l'autonomie, mais aussi la vitesse qui découle de cette autonomie.

Sur les photos on peut voir le Taranis par exemple, qui montre que la guerre est trop rapide pour être comprise par les humains qui ne peuvent ainsi réagir assez rapidement; donc nous voulons que tout aille plus vite et voulons en laisser le soin aux machines. Ceci laisse une mauvaise impression. Si quelque chose va trop vite pour que l'on puisse réagir, le problème est là, et ce problème n'est pas résolu en rendant les choses encore plus rapides, en créant encore plus de distance et en donnant aux machines encore plus d'autonomie. La personne A n'est pas contre la technologie, mais aimerait qu'elle soit conçue pour quelque chose de précis. Par exemple, les voitures sans conducteur Google: l'état doit fixer des régulations, mais nous sommes aussi bien le consommateur que l'utilisateur. Et les armes autonomes ne sont pas utilisées par le Royaume-Unis ou les Pays-Bas, mais quelque part loin de chez eux, et elles semblent tout de même effrayantes. Elles dénotent aussi une sorte de supériorité, et les armes sont souvent testées dans les régions les plus pauvres du monde. Lorsque vous parlez des avantages des armes autonomes, sont évoqués le fait de créer une distance ou de réduire les pertes dans vos propres troupes; mais si vous retournez l'argument, si ces systèmes sont utilisés contre nous, les gens auront les mêmes inquiétudes que la personne A en ce qui concerne la peur, l'absence d'imputabilité ou de prévisibilité. Les images dénotent également un côté «ce n'est pas notre problème, on ne s'en soucie pas».

Cela fait partie de la réflexivité que les humains ont au cours des guerres et que les machines n'ont pas, mais aussi de la distance que les machines créent vis-à-vis du champ de bataille. Cela implique une invincibilité que nous aimons affirmer encore plus. La question «que voulez-vous que la technologie fasse pour vous?» n'est pas assez posée vis-à-vis des armes autonomes. Souvent les états occidentaux à haute technologie prétendent qu'ils emploieront la technologie de manière convenable, mais cela n'est pas nécessairement vrai dans le futur, et cela implique que d'autres groupes ne mettront jamais les mains sur ces technologies. L'idée que nous gérons cette technologie avec prudence ne tient absolument pas compte des acteurs non-étatiques ou de la prolifération des armes autonomes.

Il y a un certain degré de déterminisme dans la discussion que la personne A n'aime pas. Même beaucoup de diplomates disent «peu importe ce que nous pensons, la technologie sera là de toute façon». Cette inévitabilité dans le débat gêne beaucoup la personne A. La conception de ces systèmes reflète également cela. Ils ont l'air si abstrait et si reluisant, et pas à la place qui leur est destiné, qui est de tuer les gens aussi rapidement que possible.

6	**Question 6. Avez-vous d'autres remarques pour cet entretien?**

Les points suivants ont été évoqués au cours de l'entretien:

La personne A a souligné que les différentes parties prenantes dans le débat devraient tous jouer leur rôle, et devraient être reconnus et estimés en tant que tel.

Un autre point sur lequel elle insiste est que nous devrions penser à ce que nous voulons que cette technologie fasse pour nous sur le champ de bataille, et y réfléchir ensemble. Un exemple est le «manuel» sur la question des drones que M. Obama a publié un peu avant sa réélection, s'étant probable-

ment rendu compte de ce que son successeur pourrait faire de ces armes.

Cette campagne amène la personne A à répéter la question: «Pourquoi?» Non parce que les personnes prennent de meilleures décisions que les machines, mais ce qu'il faut préserver est qu'il faut que ce soit un être humain qui décide de la vie et de la mort. Les humains doivent être partie prenante dans la décision de mettre à mort, et ne pas programmer cette décision dans une machine. La personne A insiste sur le fait qu'il faut que l'être humain contrôle ce que fait une machine à l'intérieur de certains paramètres. Et le débat porte aussi sur le fait de savoir ce que sont ces paramètres, et si l'on tient compte de la différence entre les systèmes offensifs et défensifs. PAX ne dit pas quels types de systèmes sont acceptables, mais les états devraient en décider. On pourrait dire que c'est à l'intérieur d'un certain environnement structuré dans lequel nous pensons qu'il est acceptable d'utiliser un système défensif afin de détruire un autre projectile. Dans l'étape suivante vers l'autonomie ces paramètres évoluent, et il faut examiner dans quelles limites ils évoluent jusqu'à l'acceptable. Une arme autonome évoluant dans une «boîte» (une zone) d'un kilomètre? Qu'en est-il de la durée? Doit-elle rester en vol 3 minutes? Pourquoi pas 10 minutes? Est-ce qu'une heure serait acceptable? Donc la question est de savoir comment l'humain peut influer sur ce qui peut être choisi. Ce débat est sur la table avec la Convention sur les Armes Conventionnelles, et le sujet n'est pas évoqué suffisamment actuellement.

J'ai évoqué la discussion que j'avais eue avec un conseiller juridique du Ministère de la Défense pour la détermination des zones de déploiement des armes autonomes, et la personne A a indiqué qu'elle ne croit pas en ce type de régulation avec ce type d'armes, car ces règles sont trop souvent transgressées, et elle trouve ce type d'arme trop effrayant; il y aura prolifération

trop rapidement. Ces «boîtes» peuvent être utilisées dans un processus diplomatique pour étudier des systèmes d'armes existants, et étudier la manière dont le contrôle humain est garanti, et en vertu de quels paramètres (spatiaux ou temporels). Et il faut voir si les systèmes actuels sont appropriés, et quel est le niveau de contrôle minimum.

Nom: Personne B
Date: 09-03-2017
Durée: 25 :26 min

No.	Question
1	**Dans quelle mesure êtes-vous partie prenante dans le débat sur les armes autonomes?**
(Question non posée du fait du manque de temps pour l'entretien (30 minutes).	
2	**Quelle est votre définition d'une arme autonome?**
Il n'existe que des définitions approximatives pour les armes autonomes, et selon les diplomates de l'ONU ceci est délibéré, car définir quelque chose est la dernière chose à faire lorsque vous voulez en obtenir l'interdiction. Pour la personne B ceci concerne les problèmes d'identification, de ciblage et de destruction de l'adversaire au sol.	
3	**Quelle est votre position vis-à-vis des armes autonomes?**
Comme pour toutes les technologies considérées comme déstabilisantes et dangereuses il devrait y avoir interdiction pour les armes autonomes avant que ne se déclenche une course aux armements et que nous franchissions un palier dans la manière de faire la guerre. Il s'agit peut-être de la troisième révolution dans l'art de la guerre. Cela constituerait un changement pro-	

fond en termes d'efficacité et de vitesse dans l'élimination de l'adversaire. Il y aura beaucoup de dommages collatéraux, et cela ne se passera pas comme les gens l'imaginent, une guerre qui serait plus froide et plus propre. Cela va potentiellement «industrialiser» la mise à mort encore plus que jamais dans l'histoire de la guerre. Ce sera terriblement déstabilisant, car dans le passé la capacité à faire la guerre dépendait de vos ressources financières (pour entretenir une grande armée et son matériel). Ces armes sont potentiellement plutôt bon marché, et on n'aura pas besoin de «main d'œuvre». Auparavant il fallait persuader des milliers de personnes de partir à la guerre, alors qu'avec les armes autonomes vous n'aurez besoin que d'un programmeur. Cela va déstabiliser l'ordre politique actuel.

Cela va probablement abaisser le seuil au-delà duquel il faut faire la guerre, et mettra plus de distance entre nous et les combats et l'acte physique de tuer. A un niveau plus moral et plus personnel, la guerre a toujours été un dernier recours, et elle doit être personnelle, sanglante et dangereuse. Quand nous prétendrons qu'elle ne l'est pas, nous combattrons probablement davantage, et cela devrait être quelque chose que les politiques doivent justifier lorsque des personnes reviennent dans des sacs mortuaires.

Mais cet argument ne s'applique pas toujours. Ces arguments sont applicables aux armes autonomes aujourd'hui, mais dans 50 ans, d'autres arguments seront peut-être plus pertinents, lorsque la technologie sera plus avancée. Pour le moment, la technologie ne peut pas être conforme aux lois humanitaires, et ne peut faire la distinction entre combattants et non combattants. Dans le futur, nous aurons des systèmes qui seront plus précis et plus en mesure de respecter les lois, mais alors d'autres arguments entrent en jeu, comme: plus les systèmes seront efficaces, plus la barrière entre la paix et la guerre sera basse.

Déployer des armes autonomes n'est pas une démarche éthique aujourd'hui, mais ce n'est pas blanc ou noir comme le problème des mines nous l'a montré. Cela amène à des lois et régulations lorsque l'on interdit l'usage, et une mine peut être considérée comme une arme autonome stupide qui ne fait pas de distinction entre les personnes.

4	**Laquelle de ces trois photos d'armes autonomes voudriez-vous discuter?** [Dans cette version du mémoire, les trois photos sont remplacées par des descriptions en raison d'éventuels problèmes de copyright]

1. Essaim de six drones
2. Sea Hunter (véhicule de surface autonome sans équipage)
3. URAN-9 (véhicule terrestre de combat à chenilles sans pilote)

5	**Pourriez-vous expliquer pour chaque photo pourquoi vous l'avez choisie, quelle valeur elle représente pour vous, et pourquoi c'est important?** [décrits]

Photo 1: de nombreux drones

Elle montre que nous parlons d'une technologie assez simple qui existe déjà aujourd'hui. Lorsqu'un essaim de drones s'approche de vous dans un contexte de guerre, vous n'aurez pas beaucoup de possibilités de défense. Il s'agira d'armes de terreur et elles seront utilisées pour terroriser des civils et des populations, et sont relativement bon marché. On peut faire beaucoup avec ces drones, qui ont l'air très simples.

Photo 2: Sea Hunter

Cette photo a été choisie, car la discussion porte sur toutes les sphères de la guerre. Ainsi, à terre, sur mer ou dans les airs, partout où l'on peut imaginer qu'une guerre se déroule, des armes autonomes seront utilisées. Voici un cas intéressant qui montre à quel point la course aux armements s'accélère. Cette

arme n'était pas connue lorsque la lettre ouverte a été publiée, mais le navire a été lancé depuis, et elle s'ajoute aux prototypes militaires d'armes autonomes. Il est intéressant de noter que lors de sa construction on a dit qu'il ne serait pas équipé d'armes, mais que son rôle serait de neutraliser les mines et de détecter les sous-marins. Mais il y a quelques mois on a admis qu'ils envisagent de le doter d'armement. Ceci montre bien que nous sommes sur une pente, et qu'il y a bien une course aux armements. Ici on n'a pas besoin de personnels ou de commodités pour faire fonctionner le navire, qui peut opérer plus longtemps. A présent on construit des navires mus par l'énergie solaire qui peuvent rester en mer pendant des années. Pas besoin d'escale pour nourrir l'équipage. On voit clairement les avantages militaires que peuvent présenter ces technologies, et les militaires seront séduits et voudront les utiliser.

Photo 3: Uran-9

Le char autonome russe montre qu'il existe beaucoup d'acteurs, la Chine, la Russie, le Royaume-Uni, Israël aussi bien que les Etats-Unis dans ce qui est clairement une course aux armements. Les armées russes ont des millions de dollars de commandes dans leurs carnets et cela représente une autre sphère du champ de bataille. On voit bien que partout où il y a des combats la guerre met en œuvre des systèmes d'armes autonomes, ce qui présente des avantages militaires évidents. Si vous regardez ce char, vous voyez qu'il n'a pas l'air trop futuriste, et que ce n'est pas ce que les gens pensent, du genre des Robots Terminator. La personne B comprend que la Russie et la Chine développent des armes autonomes quand ils savent que les Etats-Unis dépensent 18 milliards de dollars de leur budget à produire la génération future d'armes sur ce qui est fondamentalement des armes autonomes. Le problème est qu'avec cette technologie, la seule manière de se défendre est d'utiliser les armes autonomes, car personne n'a assez de temps

de réaction ou la capacité de combattre 24 heures sur 24 et 7 jours sur 7 pour se défendre, et donc là on n'a aucune défense. Par conséquent il n'est pas étonnant qu'une course aux armements ait commencé.

6	**Avez-vous d'autres remarques pour cet entretien?**

Quelques autres remarques formulées par la personne B au cours de l'entretien :

Les politiques et les militaires manquent de perspicacité quand ils pensent que l'on peut contrôler toute technologie. Nous n'avons jamais réussi à le faire avec les bombes H et A. On ne peut empêcher le génie de sortir de sa boîte.

Il sera très difficile de vérifier comment cette technologie sera mise en œuvre de manière particulière, car ici nous parlons de systèmes très complexes qui interagissent dans un environnement chaotique et complexe. Il n'est pas étonnant que l'on parle de «brouillard de la guerre»; garantir que ceci ou cela se comportera d'une certaine manière, et que l'on va pouvoir inspecter et vérifier les systèmes, est au-delà de ce qui est techniquement possible, en fait ne sera jamais possible; il pense donc qu'il est irresponsable de déployer cette technologie. Dans l'industrie, les robots peuvent être très utiles, car ils sont dans un environnement contrôlé et se débarrassent de la présence des humains, mais le dernier endroit où placer un robot est sur un champ de bataille chaotique, où l'on ne peut prévoir comment il va se comporter. On ne sait pas comment construire des «mentors» éthiques qui contrôleraient les robots, mais on doit résoudre ces problèmes pour que les voitures autonomes ne soient pas dangereuses sur nos routes. Il pense qu'un champ de bataille est le pire endroit pour résoudre ces problèmes, mais, ironie du sort, ce sera là que cette technologie fera son apparition.

La personne B aime aussi faire remarquer que les avantages provenant de l'intelligence artificielle sont énormes en termes

militaires. Les forces armées américaines ont toujours été les plus grands investisseurs dans la recherche «intelligence artificielle», et c'est une bonne chose que des personnes n'ont pas besoin de risquer leur vie ou leurs membres dans une activité de déminage. C'est un travail qui convient parfaitement à un robot, car s'il fait une erreur et qu'une mine explose personne ne va pleurer. Même chose lorsqu'on achemine des approvisionnements à travers un territoire contesté à l'aide de convois autonomes. Il existe un grand nombre d'applications de l'intelligence artificielle dans le domaine militaire qui n'impliquent pas une arme autonome prenant des décisions ultimes engageant la vie ou la mort. La personne B n'a pas trop d'objections envers les systèmes défensifs comme le Phalanx car ils travaillent dans des contextes bien définis et sont conçus pour protéger des vies et non pour les supprimer, mais évidemment toute technologie peut être réorientée. Il serait difficile de dire sur le plan moral que ces systèmes, qui existent déjà, devraient être démontés sur les navires et ainsi rendre ceux-ci plus vulnérables. Le problème se pose au niveau de l'ampleur des opérations et de la zone dans laquelle ils évoluent. C'est un cadre tout à fait différent que celui d'un drone lancé pour opérer contre des cibles 7 jours sur 7 et 24 heures sur 24 pour des opérations offensives.

La personne B se préoccupe également d'autres acteurs que nos militaires en qui nous pouvons avoir confiance, comme les terroristes, les états voyous ou des gens pour qui ce n'est pas un problème de contourner des limites éthiques, ou les pirater de manière à faire du tort. Ainsi même si nous pouvons les construire de manière à ce que ces systèmes se comportent comme nous le voulons pour combattre et que nous espérons qu'ils ne soient pas piratés, le monde sera un endroit beaucoup moins sûr.

Name: Person C
Date: 10-03-2017
Duration: 25:25

No.	Question
1	**Dans quelle mesure êtes-vous partie prenante dans le débat sur les armes autonomes?**
(Question non posée du fait du manque de temps pour l'entretien (30 minutes).	
2	**Quelle définition donneriez-vous d'une arme autonome?**
Vous avez la définition du Dictionnaire du Département de la Défense US, c'est une arme qui peut choisir et engager une cible de par elle-même fondamentalement. Ensuite on peut aller dans deux directions; pour la première, Personne C a ajouté une phrase, disant que la cible n'est pas désignée au départ par un opérateur humain. La deuxième façon de voir, et la campagne «Arrêtez les robots tueurs» s'en est de plus en plus rapprochée, est de définir les armes autonomes de manière négative, en disant essentiellement que ce sont des armes sur lesquelles les humains n'ont aucun contrôle significatif. Cela non plus ne nous aide pas beaucoup, car à un moment ou à un autre il faut définir quelque chose. Que nous définissions ce que veut dire contrôle humain significatif ou ce que signifie choisir et attaquer (une cible) n'est pas très important, mais c'est vraiment un problème difficile, car on comprend ce que cela signifie aux extrêmes, mais où se situe la ligne, là c'est compliqué. C'est où se trouve cette ligne qui détermine si ces armes sont en cours de développement ou si elles n'ont pas été déployées.	
3	**Quelle est votre position vis-à-vis des armes autonomes?**
Son opinion personnelle sur les armes autonomes est qu'il est	

difficile de prendre des décisions sur ce que nous devons faire des armes autonomes si nous ne savons même pas ce qu'elles sont. Si vous pouvez me dire ce qu'est une arme autonome et quelles sont les personnes qui sont concernées et celles qui ne le sont pas, alors je pourrai vous dire ce que j'en pense. Mais cette technologie est si récente, et les définitions sont tellement vagues qu'il est difficile d'avoir une perspective du genre «oui» ou «non» sur une chose, si la catégorie à laquelle elle appartient est potentiellement si vaste; de plus on ne sait pas dans quelle direction évolue cette technologie.	
4	**Quelles sont les trois photos d'armes autonomes que vous voudriez commenter?** [Dans cette version du mémoire, les trois photos sont remplacées par des descriptions en raison d'éventuels problèmes de copyright]
1. Véhicule terrestre à chenilles sans pilote 2. MQ-9 Reaper (Faucheuse) (Véhicule aérien sans pilote) 3. Photo rapprochée du robot de la campagne "Ban the Killer Robot" (interdire le robot tueur) devant le Big Ben à Londres.	
5	**Pouvez-vous expliquer pour chaque photo pourquoi vous l'avez choisie, quelle valeur elle représente pour vous, et pourquoi c'est important?**
1. Véhicule terrestre Celle-ci illustre le futur de la robotique militaire que les gens n'aiment vraiment pas voir ou discuter, et c'est la forme la plus courante de la robotique que l'on va rencontrer sur les champs de bataille. L'une des choses les plus intéressantes dans le débat sur les armes autonomes est que l'on assiste presque à une bataille d'analogies dans laquelle on a des groupes comme la campagne contre les robots tueurs qui qui sont axés sur des machines anthropomorphiques, comme Terminator qui entre dans une maison et essaie de déterminer si la personne qui se	

trouve à l'intérieur est un combattant ou un enfant. La Personne B et un co-auteur évoquent souvent des scénarios aériens ou navals où le «champ de bataille» est un peu plus clair, et ils présentent plus souvent des plateformes d'armes que des robots en marche.

La première photo illustre une forme essentiellement plus fonctionnelle où les robots sur le champ de bataille sont beaucoup plus probables que des Terminator qui se promènent partout. Ces systèmes sont utilisés aujourd'hui comme systèmes télécommandés.

2. MQ—9 (Reaper)

Le MQ-9 est intéressant car une grande partie des préoccupations vis-à-vis des armes autonomes ont pris naissance avec l'utilisation des drones; les gens s'inquiètent du fait que les drones rendent la guerre plus facile et que cela est dangereux. Mais on ne peut pas arrêter le développement des drones, car ils sont déjà sur le champ de bataille, et les opérations se poursuivent. Les premiers travaux ont montré que les drones proliféraient rapidement et que cela ne s'arrêterait pas, mais on peut peut-être agir sur ce qui va suivre. Et de nombreuses associations essaient d'assurer, de leur point de vue, que nous sommes parvenus à la limite; elles s'inquiètent du fait que les drones pourraient prendre des décisions par eux-mêmes.

Tout ceci touche aussi la question de la prise de décision, et quand on parle d'un robot qui prend une décision, en fait on parle d'un robot qui formule un jugement de par lui-même par opposition au robot qui adhère à sa programmation. Souvent lorsque les gens parlent des robots de manière négative, ils parlent des robots qui prennent des décisions, alors que les gens qui ont moins de préoccupations parlent des hommes qui ordonnent à des robots d'accomplir des tâches, les hommes prenant encore les décisions.

Il est fascinant de constater qu'il y a des gens qui pensent

qu'on a le droit d'être tué par un être humain, alors que les humains infligent toutes sortes de choses horribles à leurs semblables, alors que l'on sait que lorsqu'on est mort, on est mort. Il est intéressant de voir que l'on embraye aussi rapidement sur la philosophie morale, presque plus que sur tout autre chose.

3. **Robot de la campagne «robots tueurs»**

Il est intéressant de voir un groupe de personnes dont l'action a porté auparavant sur des campagnes contre les mines antipersonnel et les bombes à sous-munition et d'autres technologies qui ne sont généralement pas fondamentales dans les opérations militaires. Nombre de préoccupations vis-à-vis des armes autonomes portent sur le fait de savoir quelles armes autonomes nous avons et ce que ces armes peuvent faire, mais nous savons que l'intégration de l'autonomie dans les systèmes d'armes en général est en cours, même si la question de savoir si le système d'armes est autonome est presque une autre question; mais l'intégration de l'autonomie dans les systèmes d'armes est inévitable. Il est intéressant de constater que la campagne a choisi ce thème, en particulier parce qu'il s'agit d'un type de technologie militaire différent de celles qu'ils ont visées auparavant. Pour la Personne C on s'attend plutôt à ce qu'ils demandent l'interdiction des chars ou des sous-marins. Si la campagne contre les robots tueurs est légitime, alors les militaires voudront avoir ces robots, mais si la campagne n'est pas justifiée et que finalement ces armes ne vont pas changer grand- chose, alors le problème n'en est pas un. C'est un groupe intéressant, ces gens veulent vraiment réduire la souffrance humaine, mais ils utilisent les mêmes références pour ce qui a constitué pour eux des victoires qui étaient en fait différentes, et la Personne C pense qu'ici la situation est différente car la technologie est incertaine et la catégorie potentielle d'armement est très large, alors que personne ne connaît

l'importance potentielle de cette catégorie pour les militaires.	
6	**Avez-vous d'autres remarques pour cet entretien?**

A ma remarque qu'il n'y a pas eu jusqu'à présent beaucoup de recherche empirique, la Personne C a répondu qu'il effectuait des enquêtes par expériences d'«amorçage», ce que font en général les chercheurs qui s'intéressent à l'opinion publique, pour tester une liste de questions sur les attitudes: on présente aux gens différents contextes, et vous constatez ce que sont leurs opinions face à ces scénarios; ensuite on essaie d'évaluer ce qu'ils ressentent, mais ce travail est limité aux Etats-Unis.

Une autre chose intéressante: les stratégies sont très intéressantes, et dans une certaine mesure parce que nous vivons dans un monde dans lequel la supériorité technologique militaire américaine va de soi. Dans un monde imaginaire où la Chine déploierait des armes autonomes et où les USA et les pays européens n'en auraient pas: la semaine suivante, au Congrès américain, il y aurait des séances sur ce déséquilibre, et des questions du genre: «Pourquoi les Etats-Unis n'ont-ils pas les armes qu'a la Chine?». Ceci nous amène à la question que la Personne C évoque comme n'étant pas posée par la campagne: «Qu'arrive-t-il lorsqu'un pays non démocratique déploie ces systèmes, qui ne sont pas des armes marginales mais au contraire importantes pour les opérations militaires générales, et que ces armes vous donne une plus-value sur le champ de bataille?» «Que ferait le monde pour réagir à cela?» Ce ne sera pas nécessairement négatif, mais le problème est compliqué.

La question est de savoir s'il s'agit d'une course aux armements, dans laquelle vous êtes inquiet car quelqu'un est en train d'acquérir une technologie alors que vous êtes en concurrence directe avec lui, ou d'une prolifération rapide du fait de la facilité de l'acquisition. Une course aux armements relève de la politique car on a besoin d'une raison pour participer à cette course, et s'il y a course aux armements dans le domaine des

armes autonomes, c'est parce qu'il est difficile de savoir ce qu'est une arme autonome. Quelle est la différence entre un MQ-9 Reaper et un Reaper autonome? C'est le software et non le hardware, et vous ne le voyez pas. L'incertitude - le système du voisin est-il un drone ou une arme autonome? – constituerait un facteur essentiel qui conduirait à une course aux armements, car on ne peut pas être assuré que de l'autre côté on ne produit pas d'armes autonomes. Sauf si vous êtes branché dans leurs systèmes, mais qui vous laisserait faire cela?

Name: Person D
Date: 19-04-2017
Duration: 55:18 min

No.	Question
1	**Dans quelle mesure êtes-vous partie prenante dans le débat sur les armes autonomes?**
(Question non posée du fait que je ne l'ai pas posée aux autres personnes participant aux entretiens)	
2	**Quelle est votre définition des armes autonomes?**
C'est un système d'armes qui utilise la force, et implicitement cette force est létale, mais pas nécessairement. Ce n'est pas impératif, mais c'est une perception courante dans un contexte militaire, bien que cette hypothèse puisse être évoquée dans la discussion sur les photos. C'est du moins ce à quoi on s'attend dans les applications dans futur proche. Pour moi, l'autonomie signifie qu'elles ont la capacité de choisir elles-mêmes leur cible. De choisir un objectif dans un certain contexte, avec une mission vague. L'arme n'est pas dirigée vers des coordonnées spécifiques ou un véhicule spécifique, mais c'est plutôt «va dans cette zone et repère tous les véhicules qui se comportent de manière hostile».	
3	**Quelle est votre position vis-à-vis des armes autonomes?**
La Personne D aimerait s'y opposer, mais mon analyse montre que ce n'est pas possible; donc le mieux est de s'assurer que nous les utilisons de manière aussi responsable que possible, et donc de minimiser les risques de voir la situation échapper à tout contrôle. Je peux expliquer cette position à partir des idées suivantes: • D'un point de vue civil, il y a une forte dynamique commerciale pour développer l'intelligence artificielle en général, car c'est un marché «juteux»;	

- En même temps, dans le contexte militaire, la vitesse de prise de décision est importante. Auparavant le problème était de recueillir l'information, à présent c'est de traiter l'information. Ceci veut dire que la prise de décision humaine est en train de devenir le maillon le plus faible, ce qui conduit à penser que cette tâche sera confiée à des systèmes.
- Troisièmement il est possible qu'une intelligence supérieure à celle de l'homme soit créée. La Personne D ne voit pas de raison technique qui rendrait la chose impossible. Si c'est possible on ne peut imaginer tout ce que cela implique. Ceci, combiné à l'utilisation de la force létale, n'est pas une situation à souhaiter pour l'humanité.

A partir de ces réflexions la Personne D espère que nous ne développerons ni n'utiliserons les armes autonomes, mais cela est impossible car les deux premiers coups tirés nous mettront irrémédiablement dans cette situation. «Les plus belles fleurs sont celles qui poussent le plus près de la falaise». Ainsi, le mieux que nous puissions faire est de ne pas tomber de la falaise. Il est très pessimiste sur nos chances de pouvoir contrôler cela, mais nous devrons essayer.

Pour garder le contrôle, on pourrait utiliser un cadre, par exemple durant la phase de conception pour intégrer ces armes dans le monde que nous souhaitons avoir, et pour éviter que ces technologies créent leur propre vision des choses. Cela se situe aux premières phases du développement, mais au cours du déploiement on devra penser à la manière de concevoir le processus de ciblage. Plus ces systèmes seront capables d'agir par eux-mêmes, plus il faudra être précis sur les paramètres de ciblage, et cela pourrait être différent de ce que nous imaginons maintenant.

Nous devons essayer de parvenir à un accord international sur ce problème et c'est là le plus difficile, car ne pas respecter les règles donnera un avantage énorme. C'est le dilemme clas-

sique du prisonnier. Nous devons penser à tout ceci, ce n'est pas une question facile à résoudre, et la Personne D n'a pas de réponse à cela; mais il est évident que les parties prenantes commerciales doivent être également impliquées.

Sur la question demandant de clarifier son opposition à cette technologie: les armes autonomes sont différentes des armes conventionnelles non-autonomes en ce que ces dernières ne peuvent faire disparaître notre espèce entière, bien qu'en théorie ce soit possible avec les armes nucléaires, mais celles-ci sont trop coûteuses, alors que les armes autonomes ne le sont pas, et donc présentent un risque plus grand. Ainsi, le processus de fabrication des armes nucléaires peut être contrôlé, et il l'a été pendant environ 60 ans, ce qui nous permet d'y être habitués. Auront-nous cette possibilité avec les armes autonomes et leur rapide développement? Un autre aspect est que l'intelligence artificielle pour les armes comporte déjà un risque, contrairement à l'intelligence artificielle qui aide les médecins dans leur diagnostic, ce qui est destiné à faire le bien. Les armes autonomes combinent un système très intelligent et une intention hostile, ainsi la Personne D craint que les armes autonomes ne présentent un risque plus grand. Les armes autonomes ne sont pas la seule source de préoccupation pour la Personne D. La biologie synthétique est également un domaine à haut risque. C'est ainsi que les technologies qui ne rencontrent que des «barrières» peu élevées mais présentent des risques potentiels incontrôlables sont inquiétantes pour l'humanité.

4	**Quelles sont les trois photos concernant les armes autonomes que vous voudriez discuter?** [Dans cette version du mémoire, les trois photos sont remplacées par des descriptions en raison d'éventuels problèmes de copyright]

1. Capture d'écran du film *Ex Machina.*
2. Image d'un robot de sauvetage portant un mannequin res-

	semblant à un humain dans le cadre d'un concours de robots pour aider les spectateurs. 3. Main du robot et main d'un humain dont les index se touchent.
5	**Pouvez-vous expliquer pour chaque photo [décrits] pourquoi vous l'avez choisie, quelle valeur elle représente pour vous, et pourquoi c'est important?**
Toutes ces photos sont liées. **Photo 1.** Pour *Ex Machina*, il est bizarre de constater que la technologie qui est créée dans ce film hérite (ou assimile) également de l'instinct de survie de l'humanité, et par conséquent prend des mesures pour ne pas mourir, et aussi ne se soucie pas de ce qui arrive aux autres. Cela illustre le fait que vous devez considérer toutes les conséquences possibles lorsque vous vous lancez dans ce type de technologie, et, si vous ne faisiez pas cela, considérer ce qui pourrait mal tourner. Car nous parlons ici de systèmes qui possèdent certaines capacités à apprendre par eux-mêmes, et on pourrait ne pas voir les choses qu'ils apprennent par eux-mêmes. Il y a donc deux volets: le premier est la perte de contrôle accidentelle, et le deuxième est la recherche intentionnelle de limites, mais cela signifie aussi que vous pouvez franchir la ligne.	
Photo 2. La deuxième photo est celle d'un concours de robots secouristes sur fond de catastrophe. Le robot ne fait pas de mal, mais il ne fait rien. Alors le dilemme est que vous voulez une combinaison de 1 et 2, mais dans ce cas il faudra vraiment se rendre compte du contexte, et une sorte de «bonne volonté». La Personne D indique qu'il s'attend à ce que l'intelligence artificielle prenne de meilleures décisions que l'être humain. Il donne l'exemple d'un robot à qui on demande de prendre par écrit le plus grand nombre de notes possible, et à un moment donné il évolue pour devenir un super robot et en quelques secondes décide que la meilleure manière d'effectuer sa tâche est	

de transformer la terre entière en notes, tuer tous les humains sur terre parce qu'ils le gênent dans sa production de la plus grande quantité de notes possible. Mission accomplie, mais pas de la manière souhaitée. Il faut donc penser à cela dans la phase de conception.

Photo 3. Mais il faut également penser à la coopération entre l'homme et l'intelligence artificielle (IA); mais cela ne se produira pas dans les premières phases de sa mise en œuvre. Dans l'avenir, nous aurons des systèmes d'IA qui le feront. Les systèmes IA actuels comme le Harpy ont peu d'intelligence du contexte, mais à mesure que nous déploierons des robots dans les opérations terrestres, nous devrons coopérer avec eux, et la détermination des objectifs deviendra un point extrêmement important. La photo 3 représente cette coopération. Ceci implique que le système doit connaître nos intentions, mais nous devons aussi former nos personnels à travailler avec des systèmes de ce type, et à communiquer les objectifs à ces systèmes. Plus ces systèmes seront dotés de possibilités infinies, plus on devra penser à fixer des limites à leurs missions. La Personne D donne l'exemple: une formule magique c'est très bien, mais il faut aussi la comprendre. Ainsi, nous devons former nos personnels à donner des tâches, mais avec des limites.

6	**Avez-vous d'autres remarques pour cet entretien?**

A ma question à la Personne D lui demandant ce qu'il pense du fait qu'il doit toujours y avoir un être humain «dans le coup», il répond que cela dépend des possibilités (limitées ou illimitées) du système. D'abord, plus il aura de liberté et plus vous devrez penser, dans la phase de conception, à permettre au système de développer une «bonne volonté». En second lieu, si vous ne pouvez pas limiter ces systèmes, vous devez essayer de trouver une manière de limiter le risque.

La Personne D envisage le développement en phases. La phase 1 est une introduction à grande échelle des armes auto-

nomes ayant des capacités limitées. Pour la phase 2 on aura des équipes homme-machine dans lesquelles les personnels apprendront à coopérer avec la technologie. Mais pour la phase 3 ce sera dangereux et risqué, car nous aurons des systèmes complètement indépendants, et dans la phase 4, ce sera l'exemple de Terminator ou Skynet, qui sont des exemples dans lesquels l'IA a vraiment mal tourné. Mais la Personne D pense que tout cela n'est pas évident et ne sera pas forcément l'IA qui tourne mal, mais ce sera le cas si nous bâtissons l'IA avec le bon état d'esprit et pourtant détruisons le monde. Il pense également qu'il n'y a pas de limite bien définie entre les phases 3 et 4, mais qu'on ne pourrait s'en rendre compte que lorsqu'il sera trop tard.

Name: Person E
Date: 17-04-2017
Duration: 22:19

No.	Question
1	**Dans quelle mesure êtes-vous partie prenante dans le débat sur les armes autonomes?**
(Question non posée du fait du manque de temps pour l'entretien (30 minutes).	
2	**Quelle définition donneriez-vous d'une arme autonome?**
«Autonome» pour moi signifie que lorsque je donne l'ordre d'agir, elle trouve son chemin et choisit sa cible par elle-même (à partir des paramètres que l'homme lui a donnés), se basant sur les images de l'environnement qu'elle se crée elle-même. Donc, sélectionner, choisir et agir par elle-même constituent les trois composantes.	
3	**Quelle est votre position vis-à-vis des armes autonomes?**
Positive, je suis ouvert à ce type de technologie. Positive, mais nous devons penser à la manière de les utiliser avec circonspection, non pas les «balancer» dans l'organisation ('niet zomaar naar binnen gooien').	
4	**Quelles sont les trois photos [décrits] d'armes autonomes que vous voudriez discuter?** [Dans cette version du mémoire, les trois photos sont remplacées par des descriptions en raison d'éventuels problèmes de copyright]
1. Missile de croisière Harpoon tiré depuis un navire de la Marine. 2. Un essaim de 19 drones. 3. Robot de style futuriste brandissant une grosse arme à feu.	
5	**Pouvez-vous expliquer pour chaque photo pourquoi**

	vous l'avez choisie, quelle valeur elle représente pour vous, et pourquoi c'est important?
Photo 1.	C'est l'un des systèmes existant actuellement qui est utilisé par la Marine de Guerre depuis plus de 30 ans. C'est une torpille lancée d'un bâtiment contre les sous-marins. Dès que l'on lance cette arme elle choisit sa cible automatiquement et l'attaque. Elle n'a aucune fonction d'annulation. C'est «tire et oublie». Un système similaire est le Harpoon qui est un missile de croisière qui est utilisé contre les navires de surface. Ce type d'arme cherche sa cible et lorsqu'elle l'a trouvée, elle peut être différente de celle que vous avez ciblée. C'est un système qui est déployé avec beaucoup de prudence, mais c'est un type d'arme autonome. Du point de vue de la Navy ce système de contrôle de missile guidé peut être réglé sur un certain ensemble de paramètres, comme l'altitude, la vitesse et certaines configurations d'identification, et le système est en cours d'utilisation. Il n'y a pas d'humain dans la boucle, et nous disposons de ces systèmes depuis plus de vingt ans. Ces systèmes peuvent être utilisés sur une base autonome, mais lorsque nous les déployons nous intégrons toujours la fonction «homme dans la boucle». Un humain doit donner le feu vert avant qu'ils ne puissent engager la cible. Il existe un processus autour de cette fonction d'autonomie, et il existe même un commutateur matérialisé qui doit être actionné avant que le système ne puisse engager la cible. Une sorte de commutateur d'annulation, mais on peut régler le système de manière à pouvoir s'en passer. Ces systèmes sont développés pour être employés dans des situations où la menace aérienne est élevée, dans lesquelles chaque seconde compte. On a là déjà un certain degré d'autonomie, avec le sentiment que l'on peut intervenir à tout moment. Le fait qu'il soit possible d'intervenir à tout moment est très important pour la Navy. Ceci est tout le contraire de la torpille qui ne vous permet pas d'intervenir, ce qui veut dire

que l'on doit être absolument certain avant de l'utiliser. La Navy a développé des procédures pour le déploiement de ces deux armes.

Photo 2. Prochaine étape pour le Ministère de la Défense: les drones télépilotés que nous essayons d'intégrer dans notre organisation, mais les développements dans les technologies vont plus vite que nous en ce moment, et nous essayons de suivre le rythme. L'essaim de drones est le système que l'on utilisera dans un avenir proche. Ces systèmes seront déployés et exécuteront leur mission par eux-mêmes, mais la manière dont ils le feront nous échappe encore. Ils opéreront avec un bon niveau d'autonomie. Actuellement on s'en sert pour le recueil de renseignements et comme capteurs, mais l'étape suivante, c'est capter pour attaquer avec une arme. Cette technologie est actuellement disponible. Actuellement le Ministère travaille sur une «feuille de route» pour les systèmes aériens télépilotés, mais la barrière à franchir pour créer une pleine autonomie est trop haute. Au Ministère, cette barrière est actuellement d'ordre éthique. La question des systèmes autonomes sans pilote fait partie de la vision actuelle pour les Forces armées qui est mise sur pied par le *Hoofd Directie Beleid* (à l'EM de la Défense), et ils entament une discussion sur les armes autonomes. Le Ministère acquerra ce type d'armes après que le problème des armes autonomes aura été discuté au niveau de notre société.

Tous les spécialistes de l'Armée de Terre, de la Marine et de l'Armée de l'Air disent que l'autonomie est en marche, mais que nous voulons toujours contrôler la situation. Nous voulons être toujours capables d'arrêter les progrès lorsque la situation a changé, ou pouvoir modifier leur mission si nécessaire. Il est typique dans les opérations militaires que les situations changent rapidement, et que l'on doive adapter la planification. En particulier lorsque vous travaillez là où il n'y a pas encore de

conflit armé et que vous opérer en devant respecter les Règles d'Engagement. Vous devez pouvoir anticiper et réagir rapidement pour ne pas vous retrouver dans une situation dans laquelle vous avez déployé votre système dans l'air pour attaquer une certaine cible, et où la situation tactique ou opérative change, auquel cas il faut rappeler ces armes. Si on tient compte de cette nécessité dans la conception des systèmes, alors nous autres au Ministère de la Défense aimerions utiliser ce type de système.

Photo 3. C'est l'image que se fait le public du robot tueur, qui est un système qui ne peut pas être contrôlé et qui pense et agit de sa propre volonté. Les gens se méfient de cela; il représente la limite entre ce que nous pouvons déployer actuellement et la technologie que nous ne voulons pas adopter.

6	**Avez-vous d'autres remarques à formuler**

La Personne E a souligné qu'elle travaille depuis de nombreuses années sur les armes autonomes (comme le montre la photo 1), et qu'il existe un certain degré de contrôle qui y est intégré, tout comme il est intégré dans la gestion des procédures et opérations. En particulier lorsqu'une nouvelle technologie n'a pas fait ses preuves, il vous faudra y intégrer davantage de ces dispositifs de sécurité. La Personne E pense que nous allons dans cette direction et que nous devons le faire, car cela nous fera gagner un temps précieux; nos adversaires font la même chose, il faut donc aller aussi dans cette direction. Si un essaim de drones s'approche de vous, il vous faudra bien les contrer.

Nom: Personne F
Date: 08-05-2017
Durée: 53 :44 min

No.	Question
1	**Dans quelle mesure êtes-vous partie prenante dans le débat sur les armes autonomes?**
(Question non posée du fait du manque de temps pour l'entretien (30 minutes).	
2	**Quelle définition donneriez-vous d'une arme autonome?**
La Personne F adhère à la définition de l'ICRAC (Comité International pour le Contrôle des Armes Robotisées), qui est l'autonomie dans les fonctions critiques du ciblage et de l'engagement des cibles par une arme. Il ne s'agit pas seulement de choisir une cible comme geste fonctionnel, mais aussi d'une action anthropologique et sociologique qui est communiquée en pointant une arme sur une personne. Formellement, c'est choisir une cible ou déterminer que quelque chose est une cible valable. L'étape suivante est de lancer l'arme et faire feu avec cette arme. S'il y a autonomie dans ces deux fonctions on peut vraiment parler d'une arme autonome; si c'est pour une seule c'est plus difficile de dire si c'est une arme autonome, et jusqu'à quel point vous avez besoin d'un contrôle humain significatif. Et qu'est-ce qu'un contrôle humain significatif? Ceci fait encore l'objet d'un débat. Fondamentalement, le contrôle par l'homme, dans une perspective d'ingénierie, est un contrôle effectif, si les humains sont capables d'intervenir ou d'arrêter l'arme par un ordre, ou changer le comportement d'un système. Un contrôle non significatif, c'est lorsque la personne devient un automate borné au service d'une machine, assis dans une pièce devant un bouton, et à chaque fois qu'apparait une lumière rouge, vous être censé ap-	

puyer sur le bouton; dans un sens l'homme contrôle le bouton dans un contexte de causalité, mais ne comprend pas vraiment quelle sorte d'information est utilisée pour déterminer qu'une chose constitue une cible valable, car ce n'est qu'une lumière rouge; vous pensez que cette lumière est connectée à quelque chose qui produit des cibles valables et qui autorise à attaquer ces cibles. Mais dans ce cas on n'a aucune intelligence ou compréhension, ni aucun moyen indépendant de vérifier la légalité ou la moralité de l'attaque de cette cible. Donc effectivement on n'a pas un contrôle significatif mais un contrôle de cause à effet.

Ensuite, pour que le fait de tuer constitue un acte significatif, il doit y avoir une intention ou un objectif militaire à choisir une cible. Et simplement parce qu'une chose constitue une cible militaire légitime on doit l'attaquer. Il se pourrait que vous ne l'attaquiez pas parce que c'est cher, ou parce que vous révéleriez votre position; il existe donc d'autres considérations, même si cet objet constitue une cible militaire légitime, vous allez réfléchir pour savoir si vous allez l'attaquer ou non: voilà ce qui en fait une activité humaine; lorsque vous automatisez tout cela, cela cesse d'être significatif et devient en fait une exécution arbitraire légale. Vous n'avez aucun moyen significatif de vérifier si une cible est légitime ou significative.

3	**Quelle est votre position vis-à-vis des armes autonomes?**

La Personne F est contre, et pense que l'on devrait les interdire. Cependant on pourrait élaborer une interdiction comme exigence positive d'un contrôle humain significatif. C'est vraiment difficile d'être précis au-delà de la définition de l'ICRAC; il pense que celle-ci est plus vague d'une certaine manière que «contrôle humain significatif», ou on pourrait dire qu'il faut un contrôle humain significatif de ces deux fonctions critiques. Il pense que lorsque la notion de contrôle humain significatif sera

comprise, cela deviendra une exigence pour toutes les armes plutôt que pour une classe particulière d'arme, mais il faudra une garantie que cela sera fait pour toutes les armes. L'analogie qu'il aime faire est celle de la souffrance inutile et des blessures superflues; ce ne sont pas les mêmes définitions, mais pour elles aussi on est dans le vague sur ce que sont la souffrance inutile et les blessures superflues. Il y a eu des débats substantiels lorsque ces expressions ont été introduites à la Convention de Saint Pétersbourg et les juristes les ont adoptées; maintenant on comprend «tort supplémentaire causé qui ne sert aucune fonction militaire sauf causer la souffrance et la douleur». Même s'il s'agit de modifications ouvertes à débat, elles montrent si oui ou non nous exerçons un contrôle significatif sur le système. Est-ce que je peux l'arrêter? Est-ce que je peux intervenir? Est-ce que je peux reconsidérer les choses? Ou vérifier par moi-même que les cibles qui ont été choisies sont correctes, légalement et moralement? Car si je ne peux pas, ce système d'armes pose un problème. Les systèmes sur lesquels on a un contrôle limité sont dangereux, et au niveau mondial il faut réfléchir sur ce que l'on veut en faire.

Les systèmes autonomes actuels, comme le Goalkeeper sur des navires, devraient rester dans le cadre de ces exigences. Les états qui déploient ce genre de systèmes devrait déterminer comment ils respectent ces exigences, que ce soit pour tout système balistique ou antibalistique comme le Goalkeeper, le Phalanx ou le Patriot; car beaucoup de ces systèmes neutralisent les projectiles qui arrivent, et celui que la Personne F a vu est très limité en termes de temps et de possibilités jusqu'à ce qu'ils choisissent leur cible, moins d'une minute (20 ou 30 secondes); mais on peut toujours dire qu'il y a des hommes qui ont la main et qui décident s'ils vont laisser le système détruire sa cible. La Personne F ne tient pas à ce que l'ICRAC élimine ces types de systèmes, mais ils sont dangereux tout de même

car ils provoquent des incidents de tirs fratricides. La Personne F donne l'exemple de 2 aéronefs abattus par le système Patriot en Irak à la suite du dysfonctionnement de leur transpondeur; depuis le système Patriot a été revu au niveau de l'ingénierie; à présent l'opérateur doit donner l'autorisation formelle au système d'engager la cible et non de le laisser prendre cette décision par lui-même automatiquement. Ceci ajoute au poids psychologique qui fait que les personnels sont réticents à appuyer sur le bouton à moins qu'ils ne soient vraiment sûrs du fait de cette «opinion» de l'automate lorsque le système dit qu'il est sûr de son fait; alors les personnes consentent à laisser le système aller jusqu'au bout. Il est indiscutable qu'il y a plus de contrôle humain significatif dans la nouvelle interface Patriot qu'il y en avait dans l'ancienne interface Patriot.

4	**Quelles sont les quatre photos [décrits] d'armes autonomes que vous voudriez discuter?** [Dans cette version du mémoire, les trois photos sont remplacées par des descriptions en raison d'éventuels problèmes de copyright]

1. Squelette métallique Terminator
2. Aéronef sans pilote Predator
3. Photo du robot de la campagne "Ban the Killer Robot" devant le Parlement britannique à Londres.
4. Chaise à trois pieds devant l'ONU à Genève, en Suisse.

5	**Pourriez-vous expliquer pour chaque photo pourquoi vous l'avez choisie, quelle valeur elle représente pour vous et pourquoi c'est important?**

Photo 1. Terminator. Dans les premières années de la campagne le squelette de métal Terminator accompagnait tous les articles concernant la campagne, ce qui est intéressant d'un point de vue des médias et de l'activisme social. D'une part, c'est une image forte qui stimule la prise de conscience du public, et cela saisit son attention car il représente les Robots

Tueurs.

Terminator est très intéressant si vous vous souvenez du film, car on lui donne une cible spécifique, donc dans ce sens ce n'est pas une arme autonome, même s'il tue pas mal d'autres personnes en passant; il y a donc des décisions de ciblage autonomes, mais sa tâche première est Sarah Conner, qui est la cible réelle. Et il y a beaucoup d'autres éléments dans le film qui stimule les peurs du public, comme les robots qui deviennent conscients qu'ils existent, se retournent contre les humains et ce genre de choses; mais ce qui est plus intéressant est que cela fait penser aux armes autonomes intermédiaires et stupides qui pourraient être mises en service dans un futur entre 5 et 20 ans, qui n'auront pas de systèmes de raisonnement très sophistiqués; elles vont se baser sur un petit ensemble de critères provenant des données fournies par les capteurs, et se servir de leurs armes, ce qui peut avoir de funestes conséquences pour les civils et sur d'autres choses encore. L'image de Terminator ne fixe pas vraiment une valeur mais elle capte l'attention, et cela était important au début de la campagne. Ce n'était pas seulement la peur « dans » le film, mais une vraie peur vis à vis de la façon dont ces armes allaient opérer.

Photo 2. Predator. La deuxième chose à laquelle pensent les gens, ce sont les drones Predator et Reaper, qui sont des robots actionnés à distance, donc qui ne sont pas vraiment autonomes, et en ce sens ils ne donnent pas lieu à la même histoire. Ce sont plus de vraies images du monde qui en un sens montrent que ce sont des sortes de précurseurs des armes autonomes, dans la mesure où l'on pourrait automatiser le ciblage et l'emploi des armes des drones, et ceux-ci deviendraient des armes autonomes; il existe en réalité un développement de la prochaine génération de drones, et il y est souvent question de

ciblage autonome par exemple pour le X47-B, le Tanaris et le Neuron, qui sont des drones propulsés par réacteur et au profil furtif, qui sont équipés de beaucoup d'armes, et qui sont conçus pour évoluer dans des espaces aériens contestés, à la différence du Predator et du Reaper qui ont des hélices et sont par conséquent peu rapides, et pour l'emploi desquels il faut contrôler l'espace aérien. Dans un espace aérien où l'ennemi dispose d'aéronefs ou peu brouiller les communications, ces systèmes d'armes ne sont pas utiles. La prochaine génération de drones peut opérer dans des espaces où les communications sont interdites, ce qui est un argument en faveur de l'autonomie, car dans ce cas on n'a pas besoin de liaisons qui peuvent être interrompues et si c'est le cas on perdra le contrôle de l'aéronef; donc dans cet environnement il faut qu'il soit autonome. La question est de savoir comment il va parvenir à sa cible. Cela pourrait être comme pour les missiles de croisière auxquels on donne toute une liste de coordonnées GPS; ils trouvent ces coordonnées et les bombardent; dans ce cas ce sont les humains qui déterminent les cibles; ou alors ils vont être dotés d'une certaine capacité à choisir et sélectionner une cible par eux-mêmes parmi des objectifs inopinés, à partir des données fournies par les capteurs.

Les images de ces nouveaux systèmes ont un air sinistre, en particulier celle du Taranis, avec ces rangées de missiles sur le devant. Ceux-ci représentent ce à quoi il faut nous intéresser au sujet de la nouvelle génération des armes autonomes. Le souci serait qu'ils ont une capacité de sélection autonome des cibles. C'est un chasseur à longue portée qui choisit ses propres cibles. Ceci devient une question de légalité et de contrôle opérationnel. En fonction de la conception de ces systèmes expérimentaux, ils peuvent tomber dans la catégorie des armes autonomes ou non si les humains présélectionnent les cibles, et leur fonctionnalité n'est pas encore vraiment fixée car c'est couvert

par le secret et nous ne savons pas comment ils fonctionnent.

Le problème de fond est la dignité humaine, ce qui veut dire que la décision de tuer doit être prise par un être humain et non par une machine. Ce problème ne peut être séparé du sens que l'on donne à la mort provoquée. Un autre humain détermine que vous êtes un combattant ennemi en fonction de votre dispositif militaire sur le champ de bataille, de l'uniforme et de la participation à un conflit armé, ce genre de critères. Si c'est seulement une machine qui obéit à un algorithme, et qui essaie de suivre quelques critères qui déterminent la légalité, il n'y aura aucun jugement humain, et pour la Personne F cela signifie qu'il n'y aura aucune garantie de dignité. L'arme peut être dans le droit ou non dans une perspective légale en utilisant une fonction de probabilité, mais au bout du compte il n'y aura aucune intention ni aucun sens dans ce système (d'arme); il ne peut déterminer qu'il y a un intérêt militaire. Et il ne peut pas avoir un jugement moral ou légal, et donc s'il tue des personnes il ne peut y avoir dignité par définition. Cela peut être un acte utilitaire ou pratique pour le droit international, mais dans une perspective morale il y a absence de dignité.

A ma question de savoir si dans certaines situations les Robots prennent de meilleures décisions que les humains, du fait qu'ils ne connaissent pas la fatigue, l'émotion ou ne cherchent pas à se venger, la Personne F répond que ceci est un point important, mais que nous devons concevoir des systèmes qui aident les humains à prendre de meilleures décisions et à réduire la possibilité d'erreur, mais il faut que les humains maîtrisent les éléments qui donnent du sens à l'action et qui assurent que le processus est moral, légal et porteur de sens; les humains doivent utiliser la machine pour réduire le risque d'erreur et recevoir des suggestions documentées. Ceci aide les humains à prendre de meilleures décisions lorsqu'ils sont stressés et fatigués.

Photo 3. Robot «sympa». La troisième photo est celle de la campagne sur laquelle le robot se tient devant le Parlement britannique. C'est la photo d'un robot sympathique, la Personne F dit aimer en fait les robots, et dans les forces armées il y a des tâches qui peuvent être accomplies par des robots, mais il souligne que les tâches qui nécessitent un jugement moral et légal doivent être effectuées par des humains.

Photo 4. Chaise trois pieds représentant la campagne contre les mines anti-personnel. Une autre bonne image de la campagne est celle de la chaise à trois pieds en face du bâtiment des Nations-Unies à Genève qui symbolise la campagne contre les mines anti-personnel. Au début de la campagne pour l'interdiction des Robots Tueurs, nous avons essayé de faire la même chose que dans la campagne contre les mines, en disant qu'il s'agit d'armes qui frappent sans discrimination, avec des conséquences énormes sur les civils, ce qui est un problème. Quand on pense que les Robots Tueurs sont fondamentalement des mines anti-personnel qui peuvent se déplacer et choisir leur cible. C'est une autre manière de penser à quel point ils sont effrayants car ils ne sont pas beaucoup plus «intelligents» qu'une mine anti-personnel, et ils ne comprennent pas le monde plus que ne le fait une mine. Ils sont un peu plus sophistiqués seulement dans leur capacité de naviguer et de capter les données sur l'environnement et de sélectionner leur cible. Ainsi, d'une part ils ne seront pas aveugles au point de ne pas distinguer, et on peut même améliorer ces capacités de discrimination sur le long terme grâce à des techniques d'ingénierie, mais il reste ce bouleversement fondamental consistant à changer la nature de l'acte de tuer et ce qui justifie cet acte lorsqu'on aura ces robots tueurs automatisés. Mais ceci est vrai non seulement pour le contexte militaire et la guerre, mais aussi pour l'ensemble de la société, comme nous automatisons

un grand nombre de tâches qui étaient auparavant sous contrôle humain. Bien sûr les humains font beaucoup d'erreurs dans leurs prises de décisions, et il y a la subjectivité et toutes sortes de préjugés et autres aspects auxquels il faut remédier, mais souvent nous nous lançons dans l'automation et prétendons que la subjectivité n'existe plus du fait que l'on a affaire à des système d'ingénierie et que ceux-ci ne sont pas censés faire d'erreurs, mais ces systèmes sont compliqués et nous ne savons pas comment ils se comportent ou ce qu'ils vont faire dans des situations inattendues. Dans ce cas on soulèverait tout un ensemble de nouveaux problèmes concernant l'ingénierie et ses problèmes alors que l'on n'aurait résolu aucune question sociale.

Pour faire une autre comparaison avec les mines antipersonnel, il existe une certaine notion de prévisibilité mais aussi apparemment d'imprévisibilité. Les mines anti personnel peuvent se justifier pour effectuer un coup d'arrêt au cours d'une guerre mais 10 années plus tard, lorsque quelqu'un veut cultiver la terre, les mines peuvent encore exploser, et là on ne peut contrôler ce qui se passera dans le futur. Ceci constitue le problème intrinsèque des armes autonomes dans la mesure où vous savez où elles sont censées se diriger, trouver leur cible et les détruire. Cela semble raisonnable, seulement si elles ne peuvent capter qu'un signal radar spécifique, comment peut-on être sûr de cela? Combien de choses ressemblent à ce signal radar? Il y a d'autres objets qui émettent des signaux et qui pourraient générer une confusion, ou des individus pourraient placer des émetteurs imitant ces signaux radar pour attirer le missile sur d'autres objets. Si on le maintient en opération pendant longtemps ou sur une vaste zone, on ne saura pas vraiment ce que l'arme autonome va faire. Vous savez ce qu'elle est censée faire, mais après vous ne pourrez pas vraiment la contrôler. Et c'est là que les choses deviennent dangereuses, car toute la

question est celle du degré de risque, et on ne sait pas vraiment mesurer cela, et comme ces systèmes deviennent de plus en plus compliqués, cela devient de plus en plus difficile de le mesurer, à la fois pour l'opérateur qui essaie de prévoir ce qu'il faut faire, et pour le tribunal si on essaie de trouver la personne responsable d'avoir libéré quelque chose de terrible plutôt que d'avoir l'intention de faire quelque chose de terrible. Donc on ne peut pas vraiment accuser quiconque de crime de guerre, mais de négligence puisqu'on ne savait pas ce que l'arme allait faire du fait d'une circonstance survenue de manière inattendue ou dont nous n'avons pas eu connaissance. Ainsi, tout notre système d'explication de la responsabilité et d'imputabilité commence à s'écrouler, et cela crée un grand vide pour des individus qui veulent le mal et ne pas être tenus pour responsables; cela place aussi un fardeau sur les épaules des conscrits ou de ceux qui s'engagent dans l'armée lorsqu'ils ont 18 ans et qui ne sauront pas ce qu'ils vont faire ou de quoi ils seront tenus pour responsables.

6	**Question 6. Avez-vous d'autres remarques pour cet entretien?**

J'ai posé à la Personne F la question suivante: pourriez-vous imposer une interdiction sur ce type de technologie? Réponse: Il y a des choses qu'il ne faut pas oublier concernant le droit international et les interdictions. L'imposition est l'un des éléments, mais il y en a d'autres. On peut comparer cela au meurtre. Les gens continueront à s'entretuer, mais alors on n'aurait pas dû proscrire le meurtre? On le proscrit car c'est une mauvaise action, et les gens continuent à la commettre, et il faut les arrêter. La question est d'établir une norme à l'échelle internationale, rassembler les pays du monde entier pour qu'ils tombent d'accord sur le fait que fondamentalement c'est mal de posséder un système d'armes autonome. A partir de là, quelles seraient les implications si des pays ne respectaient pas

l'accord? La plupart des traités n'ont pas de méthode de vérification ni de sanctions explicites ou ce genre de choses, mais il faudra jeter l'opprobre sur ces pays, les sanctionner et leur faire subir les conséquences.

Pour les acteurs non-étatiques, c'est une autre planète, et déjà nous avons fait interdire les armes chimiques; c'est une bonne analogie car les pays, même ceux qui n'ont pas signé le traité, sont pointés du doigt et sanctionnés, et subissent les conséquences s'ils utilisent des armes chimiques. Cela affecte aussi l'industrie dans le sens que les grands entrepreneurs traitant avec l'armée ne développent pas activement des armes chimiques, et les grandes entreprises de produits chimiques ne produisent pas d'énormes quantités d'agents précurseurs qui sont interdits par la convention. On va assister à quelque chose de semblable pour les armes autonomes. Evidemment on peut construire une arme autonome à partir d'éléments que l'on peut se procurer dans un magasin de bricolage, mais ce ne sera pas un système militarisé qui peut détruire des villes et des usines. Et déjà des individus peuvent construire des pièges, des bombes et ce genre de choses; on ne peut pas arrêter cela, mais on essaie de réguler l'accès aux explosifs ou aux technologies clés. On n'empêche les armes autonomes de proliférer et de migrer des grands états aux petits, qui eux les utiliseront sans restrictions ou sans modifier leurs logiciels, et finalement de tomber entre les mains d'acteurs non-étatiques ou de terroristes qui auraient accès à des systèmes ayant de sérieuses capacités, et qui ne seraient pas seulement des jouets auxquels on aurait attaché une bombe. Donc, avec un traité international on peut remédier à cela. Mais rien n'empêche des forces de police ou des états de les utiliser contre leurs propres peuples; donc il nous faut des lois et des normes nationales pour également empêcher l'emploi de ces systèmes par des forces de police contre des manifestants. Si des particuliers construisaient de

tels systèmes, cela tomberait aussi sous le coup du droit civil et pénal; mais ce serait une bonne chose de préciser qu'il est interdit de fixer des armes à des systèmes de ciblage autonomes.

J'ai ensuite demandé à la personne F si cela impliquerait des critères de transparence. Réponse: si la vérification des systèmes militaires sophistiqués se fait sérieusement, prouver qu'il existe un contrôle humain significatif implique prouver qu'il existe une chaine de commandement et des bases de données; si quelqu'un croit que votre système en opération est pleinement autonome, on peut prouver grâce à une suite de données qu'en réalité la décision au niveau du ciblage est prise par un humain, ou que l'être humain a autorisé le choix de la cible dans ce cas précis, ceci pour tous les types de systèmes d'armes. Il est facile de disposer d'un système de traitement de l'information efficace, mais disposer de l'autorité pour y avoir droit de regard, c'est autre chose. La norme actuelle est que l'on impose cette discipline aux militaires; dans ce cas il n'est pas nécessaire de vérifier, mais il faut qu'il existe des procédures permettant de faire passer les officiers et les soldats en cour martiale s'ils violent les lois internationales. On ne s'attend pas à ce que les ennemis organisent un tribunal pour cela. On vérifie ses propres armes, ainsi ce n'est pas du tout transparent, et on ne publie pas ces vérifications. Mais l'exigence serait que vous conservez les données, et si un événement survient vous aurez les données disponibles.

J'ai ensuite demandé s'il a eu vent des dernières informations concernant la campagne, qui a été très suivie en 2015 et 2016, mais qui a l'air de s'essouffler. Réponse: le problème est que du fait de la durée entre les réunions de l'ONU et la publication des nouvelles (en décembre) les discussions sont passées des réunions informelles entre experts au Groupe Intergouvernemental d'Experts pour les Armes Autonome Létales (Group of Governmental Experts on Lethal Autonomous Weapon Sys-

tems: GGE (LAWS)), ce qui constitue la dernière phase avant la négociation du traité. Ils ont prévu deux semaines pour ces réunions, mais ils ont eu pas mal de problèmes de financement cette année car la budgétisation a changé à l'ONU qui demande que les sommes dues soient payées à l'avance pour que vous teniez ces réunions, alors que ces 50 dernières années c'était un système différent en ce sens qu'ils tenaient d'abord les réunions, puis ils attendaient que les gens paient ce qu'ils doivent plus tard. C'est ainsi qu'on se demande si la deuxième semaine ces réunions auront vraiment lieu si les pays ne paient pas ce qu'ils doivent intégralement. La réunion du mois d'août va continuer et nous espérons qu'il y en aura une autre en novembre, en marge de leur réunion annuelle. Ils ont bon espoir que quelque chose va sortir du GGE, mais une semaine ce n'est vraiment pas assez long; ils espéraient trois semaines de réunions et non pas 2 qui pourraient être réduites à une seule.

En général, la campagne a démarré «sur les chapeaux de roues» en comparaison de la plupart des campagnes sur le désarmement, y compris celles sur les mines anti-personnel et les bombes à sous-munitions pour lesquelles il a fallu 10 ans pour en arriver là où nous en sommes au bout de trois ou quatre ans. Maintenir ce rythme eût été impressionnant, et ce qu'il l'a freiné fut que les états étaient intéressés par la discussion, mais beaucoup moins pour formaliser leur position. Alors ils paralysent plus ou moins les choses, mais il est certain que les Nations-Unies formalisent leur position, en particulier avec la politique 3000/09 étudiée avec le Pentagone actuellement, qui a été initiée il y a cinq ans. Et l'ONU espère que les états discuteront davantage dans le cadre du GGE, et accorderont leur langage, car il est facile de dire qu'il n'y a pas de définition pour l'autonomie et les armes autonomes, et nous ne savons pas ce que signifie «contrôle humain significatif»; c'est à eux de décider qu'ils peuvent justifier tous ces termes comme ils le veu-

lent. Mais pour un traité il faut trouver un consensus politique sur ces définitions. L'ICRAC fait du bon travail en restreignant la définition pour «arme autonome»; ensuite la question est de définir le «contrôle humain significatif», et quel langage adopter pour un traité qui ne serait pas trop permissif ni trop restrictif. Cet équilibre est difficile à trouver; c'est un problème politique plus qu'un problème technique. La Personne F espère que l'on va faire quelques progrès.

Une des raisons pour lesquelles le CCW (pour les armes conventionnelles) était si impatient était qu'ils n'ont pas fait grand-chose ces dix dernières années, et ce qu'ils avaient fait était insuffisant; donc ils essaient de se racheter, et c'est une bonne chose. Si nous y parvenons bientôt ce sera surtout un traité préventif, mais les armes existant actuellement doivent être révisées de par ces nouvelles exigences concernant le contrôle humain significatif; cela devrait être possible pour la plupart des systèmes, et ceux qui n'offrent pas cette possibilité ne devront pas être utilisés de toute manière que ce soit. Ce sera plus facile de le faire maintenant de façon préventive; dans dix ans les états auront développé des drones et des sous-marins pleinement autonomes, et jugeront que ceux-ci sont nécessaires à leur sécurité, et alors il sera impossible de faire adopter ce traité.

Appendice G. Mémos d'encodage

Cahier d'encodage chercheur 1

Dans l'*encodage des valeurs* le texte est encodé pour trouver les valeurs,[11] les attitudes[12] et les convictions[13] (Saldana, 2015). Ces concepts sont listés avec des codes préétablis.

Liste des concepts préétablis:

Concept	**Couleur**
Valeur	Bleu
Conviction	Vert
Attitude	Jaune

Liste des codes émergents:

Concept	**Couleur**
Définition	Violet

RECOMMANDATIONS:

- Une liste préétablie peut comporter entre 10 et 40 ou 50 codes. Nous conseillons de ne pas créer trop de codes, car la personne qui encode peut se trouver débordée ou faire des erreurs dans le processus d'encodage s'il y en a trop.

[11] Valeur: l'importance que l'on attribue à soi-même, à une autre personne, objet ou idée.

[12] Attitude: la manière dont nous pensons et nous sentons vis-à-vis de nous-mêmes, d'une autre personne, objet ou idée.

[13] Conviction (opinion): fait partie d'un ensemble qui comprend nos valeurs et attitudes plus notre connaissance, expérience, opinions, préjugés personnels et autres perceptions interprétatives de l'environnement social.

- La règle de base pour l'encodage est de faire en sorte que les codes correspondent aux données, plutôt que de faire en sorte que les données correspondent aux codes.
- Créer des mémos au cours du processus d'encodage fait partie intégrante des approches d'encodage, qu'elles soient fondées ou faites à priori. La recherche qualitative est par essence réflexive; comme le chercheur approfondit son sujet, il est important qu'il enregistre son processus de réflexion dans des mémos de nature réflexive ou méthodologique; ceci peut permettre de mettre en lumière ses propres interprétations subjectives des données. Il est crucial de commencer ces mémos dès le début de la recherche. Indépendamment du type de mémo produit, ce qui est important est que le processus initie une production et une réflexion critique dans la recherche. Ceci facilitera les analyses et les rendra plus cohérentes à mesure que le projet avance. Les mémos peuvent être utilisé pour planifier les activités de recherche, découvrir du sens dans les données, maintenir le rythme de progression de la recherche et l'assiduité, et rendre possible la communication.

Plus d'info:
http://onlineqda.hud.ac.uk/Intro_QDA/how_what_to_code.php
https://researchrundowns.files.wordpress.com/2009/07/rrqualcodinganalysis_7_19_09.pdf
http://programeval.ucdavis.edu/documents/Tips_Tools_18_2012.pdf
https://www.slideshare.net/kontorphilip/qualitative-analysis-coding-and-categorizing

NOTES:
Avant de commencer:

Mon intention est d'utiliser la méthode d'encodage des valeurs (Saldana, 2015); mais Saldana (2015) donne une définition différente pour «valeur» de celle de Friedman et Kahn Jr (2003). Ainsi, je me demande si la différence entre valeur, attitude et conviction est vraiment claire, mais je verrai ce qu'il advient. Peut-être un nouveau nom de code pour une «définition» doit être ajouté.

En cours d'encodage:

a. Je remarque qu'une attitude implique souvent un verbe; du moins c'est comme cela que je l'interprète.

b. Une conviction implique souvent une opinion.

c. Je remarque que je pense souvent aux listes des valeurs de mon étude de documents lorsque je souligne un terme, par exemple «réflexivité». Ceci pourrait indiquer une subjectivité de ma part en tant que chercheur; il serait bon de vérifier ma liste de valeurs par rapport à celle du deuxième réviseur.

d. Si une personne interviewée n'aime pas tel ou tel terme, comme «déterminisme», je l'ai mis en relief en tant que valeur.

e. Est-ce qu'une chose souvent évoquée est importante pour une personne interviewée? Et donc pourrait-elle être interprétée comme étant une valeur? Par exemple «distance entre l'homme et la machine». (Voir j.).

f. En cours d'encodage, je suis souvent revenue sur mes pas pour voir comment j'avais classifié la même expression (par exemple «trop effrayant») et recopié l'encodage.

g. J'ai ajouté un nouveau code pour «définitions», et l'ai ajouté au cahier d'encodage.

h. J'ai interprété une opinion de la personne interviewée comme «conviction», car cela fait partie de la définition du terme «croire» donnée par Saldana (2015).

i. Après l'encodage d'une question tirée d'un entretien, j'ai relu ma question et j'ai vérifié l'encodage.

j. Si un terme a été utilisé plus que deux fois, par exemple le terme «défensif» au cours de l'entretien avec le Professeur Walsh, je l'ai mis en relief en tant que valeur.
k. J'ai également mis en relief un terme qui était spécifiquement souligné, par exemple «gain / intérêt» («benefit») était marqué comme valeur dans l'entretien avec le Professeur Walsh.
l. Après avoir encodé un nouvel entretien, j'ai vérifié le(s) précédent(s) pour voir si j'étais cohérente.
m. Si quelque chose était marqué comme important, par exemple dans l'entretien avec Ad sur l'importance pour la Marine d'avoir la capacité d'intervenir n'importe quand, je l'ai encodé comme valeur.
n. Après avoir encodé les trois derniers entretiens, je n'ai pas revérifié pour les trois premiers. Du fait surtout des restrictions d'horaire, mais aussi parce que l'encodage devenait plus facile pour moi.

Après l'encodage:
J'ai envoyé tous les 6 entretiens au deuxième réviseur, et j'ai ajouté quelques informations générales sur l'encodage des entretiens. J'ai inclus le cahier d'encodage sans mes remarques. Nous avons discuté de ce travail via Skype pour s'assurer qu'elle comprenait.

Cahier d'encodage chercheur 2
J'indique ici les orientations établies à partir des hypothèses, des principes ou des raisonnements qui ont été utilisés pour diriger les propos vers les catégories valeur, attitude ou conviction. Il faut noter que ces listes se sont développées et élargies au cours de ce processus d'encodage, et donc le processus est devenu itératif. En d'autres termes, dans les derniers entretiens, j'ai toujours réfléchi sur les orientations qui facilitaient le pro-

cessus d'encodage dans les entretiens précédents, et j'ai progressé à partir de cela.

Valeur: lorsque des propos reflètent un fort degré d'importance, comme «nous devons», «il faut que», «important».

- Lorsqu'ils reflètent un jugement fort ou de haute portée, par exemple «assurer», «inquiétant», «besoin», «espèce humaine».
- Lorsqu'ils contiennent des mots d'emphase ou grandiloquents, par exemple «assurer», «souffrance humaine», «inévitable».
- Lorsqu'ils reflètent une chose considérée d'importance ou ambitieuse, «besoins (exigences)», «réflexivité».
- Lorsqu'ils constituent une partie significative d'une histoire ou d'une discussion, ou illustre quelque chose d'important, par exemple «exigence», «moralité».
- Lorsqu'ils illustrent une chose jugée digne d'intérêt, par exemple «justifier», «droit humanitaire».

Attitude: Lorsque les propos dénotent des tendances ou une direction, par exemple «non seulement», «trop».

- Lorsqu'il s'agit d'opinion ou de préférence personnelles, par exemple «vague», «pas possible», «vous voulez».
- Lorsqu'il s'agit d'une position relative, par exemple «n'a pas d'importance», «vraiment difficile», «vague».
- Lorsqu'il s'agit d'une chose qui pourrait affecter le jugement, par exemple «pas pour», «plus que», «nous voulons».
- Lorsqu'on emploie des mots illustrant des points de vue ou qui sont répétés, comme «contre», «vague», «terrible».

- Lorsque les mots reflètent une certaine prise de position, comme «fait délibérément», «devrait être», «terriblement», «séduit».

Conviction: lorsque les propos sont basés sur des hypothèses, ou concernent une chose considérée comme vraie

- Lorsqu'ils sont basés sur des attentes ou des hypothèses, comme «nous pourrions», «cela sera».
- Lorsqu'on discute de données inconnues ou des perspectives supposées d'autres personnes.
- Lorsque les propos reflètent des connaissances / expériences personnelles, comme «survient / arrive» ou «ne sais pas».
- Lorsqu'ils concernent des questions, spéculations ou des hypothèses, comme «ils n'ont pas», «sera».
- Lorsqu'ils concernent des perceptions interprétatives ou futures, comme «il n'y a que», «sera».

Appendice H. Résultats processus d'encodage

Entretiens		**Valeurs**	
	Chercheur	*Second réviseur*	*Valeurs similaires*
Entretien 1 (Personne E)	• Intervenir n'importe quand • Contrôler • Les rappeler • Ne peut être contrôlé • Niveau de contrôle	• Nous devons réfléchir à manière de les utiliser ai avec prudence. • Le fait qu'il doit être possible d'intervenir n'importe quand est très important pour la Marine. Mais nous voulons toujours contrôler. • Nous voulons toujours pouvoir les rappeler lorsque la situation a changé. • Ou pouvoir redéfinir leur mission si nécessaire. • On doit pouvoir les rappeler. • On doit intégrer davantage de ces systèmes de sécurité. • Nous devons progresser.	• Intervenir n'importe quand. • Contrôler. • Les rappeler.
Entretien 2 (Personne D)	• Echappe au contrôle. • Contrôle. • Contrôle. • Risque.	• S'assurer que nous les utilisons de manière aussi responsable que possible. • Minimiser le risque de voir la situation échapper à tout contrôle.	• Minimiser le risque. • Contrôler. • Bonne volonté. • Fixer des

	• Contrôlé. • Risque. • Faire bien. • Risque. • Risque incontrôlable. • Echappe au contrôle non intentionnellement. • Limites. • Bonne volonté • Fixer des limites. • Missions avec limitations. • Limites. • Sans limites. • Bonne volonté. • Risque. • Risqué.	• Dans le contexte militaire la promptitude de la prise de décision est importante. • Ce n'est pas une situation que nous souhaitons pour l'espèce humaine. • S'assurer que nous ne tombons pas de la falaise. • Empêcher que ces technologies créent leur propre vision. • Les technologies à barrières peu élevées, mais potentiellement risquées et incontrôlables sont inquiétantes pour l'humanité. • Considérer les conséquences possibles lorsque l'on se lance dans ce type de technologie, et ce qui pourrait mal tourner si tel n'était pas le cas. • Nous aurons besoin d'une intelligence du contexte développée ainsi qu'une sorte de «bonne volonté». • Mission accomplie mais pas de la manière dont nous l'avions voulu. • Devrons coopérer et la détermination des objectifs deviendra extrêmement importante. • Les systèmes doivent connaître nos intentions.	limites

		• Nous devons former nos personnels à travailler avec ces systèmes, et à communiquer les objectifs à ces systèmes. • Il faut former nos personnels à assigner des missions avec des limitations. • Bonne volonté. • Restreindre le risque. • Construire l'intelligence artificielle avec le bon état d'esprit mais détruire le monde quand même.	
Entretien 3 (Personne C)	• Décisions. • Décider • Forme fonctionnelle • Décisions • Décision • Décision • Décisions • Décision • On a le droit d'être tué par une personne • Souffrance humaine	• Comment être sûr que c'est là la limite • On a le droit d'être tué [par une personne] • L'intégration de l'autonomie dans les armes est inévitable • Réduire la souffrance humaine • Ces stratégies sont très intéressantes • Vous donne l'avantage sur le champ de bataille ? • On ne peut pas être assuré que de l'autre côté on ne construit pas des armes autonomes.	• Droit à être tué par une personne • Souffrance humaine
Entretien 4 (Personne A)	• Cadre éthique. • On ne peut pas laisser le contrôle	• Une arme sans contrôle humain dans les fonctions critiques, comme le ciblage, doit être interdite. • Contrôle humain signifi-	• Contrôle humain significatif. • Responsabilité hu-

	nous échapper. • Contrôle humain. • Distance. • Contrôle humain significatif. • Responsabilité humaine. • Contrôle humain significatif. • Distance. • Distance. • Distance. • Distance. • Absence de responsabilité [et] imprévisibilité. • On ne s'en soucie pas. • Réflexivité. • Distance. • Invincibilité. • Déterminisme. • Inévitabilité. • Besoin humain d'être impliqué dans la décision	catif. • Responsabilité humaine. • Le contrôle humain doit être appliqué au moment du ciblage. • Les systèmes d'armes actuels doivent être révisés concernant l'intégration du contrôle humain. • Délibérément conçu pour quelque chose. • Créer une distance ou réduire le tort causé. • Peur, absence de responsabilité et imprévisibilité. • Réflexivité. • Distance. • Invincibilité. • Déterminisme. • Inévitabilité. • Besoin humain d'être impliqué dans la décision de supprimer des vies. • Besoin qu'ont les humains de contrôler une machine à l'intérieur de certains paramètres. • Comment le contrôle humain est garanti.	maine. • Prévisibilité. • Réflexivité. • Distance. • Invincibilité. • Déterminisme. • Inévitabilité. • Besoin humain d'être impliqué dans la décision de supprimer des vies.

	de supprimer des vies. • Besoin qu'ont les humains de contrôler une machine.		
Entretien 4 (Personne A)	• Cadre éthique. • On ne peut pas laisser le contrôle nous échapper. • Contrôle humain. • Distance. • Contrôle humain significatif. • Responsabilité humaine. • Contrôle humain significatif. • Distance. • Distance. • Distance. • Distance. • Absence de responsabilité [et] imprévisibilité. • On ne s'en	• Une arme sans contrôle humain dans les fonctions critiques, comme le ciblage, doit être interdite. • Contrôle humain significatif. • Responsabilité humaine. • Le contrôle humain doit être appliqué au moment du ciblage. • Les systèmes d'armes actuels doivent être révisés concernant l'intégration du contrôle humain. • Délibérément conçu pour quelque chose. • Créer une distance ou réduire le tort causé. • Peur, absence de responsabilité et imprévisibilité. • Réflexivité. • Distance. • Invincibilité. • Déterminisme. • Inévitabilité. • Besoin humain d'être impliqué dans la décision de supprimer des vies. • Besoin qu'ont les hu-	• Contrôle humain significatif. • Responsabilité humaine. • Prévisibilité. • Réflexivité. • Distance. • Invincibilité. • Déterminisme. • Inévitabilité. • Besoin humain d'être impliqué dans la décision de supprimer des vies.

	soucie pas. • Réflexivité. • Distance. • Invincibilité. • Déterminisme. • Inévitabilité. • Besoin humain d'être impliqué dans la décision de supprimer des vies. • Besoin qu'ont les humains de contrôler une machine.	mains de contrôler une machine à l'intérieur de certains paramètres. • Comment le contrôle humain est garanti.	
Entretien 5 (Personne F)	• Contrôle humain significatif. • Contrôle humain significatif. • Contrôle significatif. • Intention. • Significatif. • Significatif. • Significatif. • Contrôle humain significatif.	• Contrôle humain effectif. • Intelligence ou compréhension contextuelle ou situationnelle. • Vérification de la légalité ou moralité. • Intention ou objectif militaire dans le choix d'une cible. • Contrôle humain significatif • Souffrance non nécessaire et blessures superflues.	• Contrôle humain significatif. • Souffrance non nécessaire. • Blessures superflues. • Dignité humaine. • Suppression de la vie dans la dignité. • Prévisibili-

	• Contrôle humain significatif. • Contrôle humain significatif. • Contrôle humain significatif. • Souffrance inutile. • Souffrance non nécessaire. • Blessures superflues. • Contrôle significatif. • Subjectivité automate. • Contrôle humain significatif. • Contrôle. • Dignité humaine. • Degré du significatif. • Digne. • Dignité. • Intention. • Sens. • Digne. • Significatif. • Contrôle humain.	• Valide, moralement et légalement. • Systèmes intelligents sophistiqués. • Perte de contrôle. • Légalité et contrôle opérationnel. • Dignité humaine. • Action de tuer légale et morale. • Tuer dans la dignité. • Besoin militaire. • Jugement moral ou légal. • Assurer un processus moral, légal et significatif. • Prévisibilité. • Contrôle. • Responsable. • Imputabilité et responsabilité. • Lois et normes. • Chaîne de commandement. • Autorité. • Transparent. • Conserver ces données.	té. • Contrôle. • Imputabilité. • Responsabilité.

	• Subjectivité. • Subjectivités. • Prévisibilité. • Non prévisibilité. • Contrôle. • Contrôle. • Responsable. • Imputabilité. • Responsabilité. • Imputable. • Responsable. • Contrôle humain significatif. • Responsable. • Contrôle humain significatif • Contrôle humain significatif • Contrôle humain significatif.		
Entretien 6 Personne B)	• Défense. • Qui défend. • Défend.	• Justifier. • Droit humanitaire. • Efficacité. • Barrières.	• Défense. • Tort causé.

	• Défend. • Irresponsable. • Avantages. • Sans défense. • Tort causé.	• Lois et régulations. • Défense contre ceci. • Avantages militaires. • En vous défendant. • Précision. • Vérifier. • Se comporter. • Se comporter d'une certaine façon. • Audit et vérification. • Environnement. • Comportement attendu. • Gouvernance éthique. • Enveloppes bien définies. • Sauvegardes éthiques. • Tort causé. • Se comporter comme on le voudrait. • Ne peut être piraté.	

Basée sur les entretiens et la comparaison de l'encodage des valeurs du chercheur et du deuxième réviseur, la liste suivante de 23 valeurs uniques peut être établie:

1. Intervenir à tout moment
2. Contrôler
3. Les rappeler
4. Minimiser le risque
5. Contrôler
6. Bonne volonté
7. Fixer les limites
8. Droit à être tué par une personne
9. Cadre éthique
10. Contrôle humain significatif

11. Prévisibilité
12. Réflexivité
13. Distance
14. Invincibilité
15. Déterminisme
16. Inévitabilité
17. Souffrance non nécessaire
18. Blessures superflues
19. Dignité humaine
20. Imputabilité
21. Responsabilité
22. Défense
23. Tort causé

Ces cinq dernières années, un débat sur le déploiement d'armes autonomes est poursuivi dans la société et à la Convention sur certaines armes classiques (CCAC) aux Nations Unies. Cependant, peu de recherches empiriques ont été menées qui permettent de comprendre comment les armes autonomes sont perçues par le grand public et les militaires. Dans cet ouvrage, nous décrivons une étude empirique qui permet de comprendre 1) comment les armes autonomes sont perçues par les militaires et le grand public, et 2) quelles sont les valeurs morales que les gens considèrent comme importantes lorsque des armes autonomes seront déployées dans un avenir proche. Nous avons utilisé la méthode de conception sensible aux valeurs *(Value-Sensitive Design, VSD)* dans notre étude comme approche de recherche. La VSD est une approche en trois parties qui permet de prendre en compte les valeurs humaines tout au long du processus de conception de la technologie. Il s'agit d'un processus itératif pour l'investigation *conceptuelle, empirique* et *technologique* des valeurs humaines impliquées par la conception. Nos résultats indiquent que les militaires et les civils travaillant au Ministère de la Défense néerlandais attribuent plus d'agentivité (qui est la capacité de penser et de planifier) à une arme autonome qu'à un drone commandé par l'homme. Nous avons également constaté qu'il existe un terrain de convergence entre les militaires et les groupes sociétaux dans le débat sur les armes autonomes, notamment en ce qui concerne la perception des valeurs que sont la dignité humaine et l'anxiété. Ces deux valeurs sont souvent mentionnées dans le discours, il serait donc essentiel de les aborder lors du débat sur l'éthique du déploiement des armes autonomes.

Ce mémoire a reçu le premier prix de l'année 2018 dans le cadre du concours annuel d'EuroISME pour le meilleur mémoire d'étudiant (Master of Arts). Pour plus d'informations sur le concours, veuillez consulter le site www.euroisme.eu

EURO ISME
International Society for Military Ethics in Europe
Awarded Special Prize "Europe of Peace" of the European Award for Citizenship, Security and Defence 2013