L'énigme de la toxine

Gregory Kavka

Vous vous sentez extrêmement chanceux. Un milliardaire excentrique vient de vous aborder en vous proposant le marché suivant. Il place devant vous une fiole de toxine qui, si vous la buvez, vous rendra terriblement malade pendant une journée, mais ne menacera pas votre vie et n'aura aucun effet durable (votre épouse, une as de la biochimie, confirme les propriétés de la toxine). Le milliardaire vous paiera un million de dollars demain matin si, à minuit ce soir, vous avez *l'intention* de boire la toxine demain après-midi. Il souligne que vous n'avez pas besoin de boire la toxine pour recevoir l'argent ; de fait, si vous réussissez, l'argent sera déjà sur votre compte bancaire des heures avant qu'il soit temps de le boire (ceci est confirmé par votre fille, une avocate, qui a auparavant examiné les documents juridiques et financiers que le milliardaire a signés). Tout ce que vous avez à faire est de signer l'accord puis d'avoir l'intention, à minuit ce soir, de boire la substance demain après-midi. Vous êtes parfaitement libre de changer d'avis après avoir reçu l'argent et donc de ne pas boire la toxine (la présence ou l'absence d'intention sera déterminée par le scanner cérébral et dispositif informatique à « lecture d'esprit » dernier cri du grand Docteur X. En tant qu'expert en sciences cognitives, matérialiste et ancien élève fidèle du Docteur X, vous n'avez aucun doute que la machine détectera correctement la présence ou l'absence de l'intention en question).

Face à cette offre, vous signez avec empressement le contrat en pensant « voilà un moyen facile de devenir millionnaire ». Peu de temps après, néanmoins, vous commencez à vous inquiéter. Vous pensiez pouvoir éviter de boire la toxine et simplement empocher le million. Mais vous réalisez que si vous pensez en ces termes lorsque minuit sonnera, vous n'aurez pas l'intention de boire la toxine demain. Donc peut-être devrez vous vraiment boire la substance pour ramasser l'argent. Ce ne sera pas agréable, mais devenir millionnaire vaut sans aucun doute une journée de souffrance.

Néanmoins, comme vous l'aurez immédiatement remarqué, il n'est pas vraiment nécessaire de boire la toxine pour empocher l'argent. Cet argent sera ou ne sera pas sur votre compte bancaire à 10 heures demain, vous saurez alors s'il y est ou pas, et que vous buviez ou que vous ne buviez pas la toxine des heures plus tard ne pourra pas avoir d'effet sur la transaction financière réalisée. Donc au lieu de planifier l'absorption de la toxine, vous décidez d'avoir l'intention aujourd'hui de la boire puis de changer d'avis après minuit. Mais si c'est ce que vous avez prévu, il est évident que vous n'avez pas l'intention de boire la

toxine (au mieux vous avez l'intention d'avoir l'intention de la boire.) Avoir une telle intention est incompatible avec le fait de prévoir de changer d'avis demain matin.

C'est alors que votre fils, un stratège au Pentagone, fait une suggestion utile. Pourquoi ne pas vous obliger à boire la substance demain, en faisant aujourd'hui des arrangements irréversibles qui vous donneront une motivation suffisante et indépendante de la boire ? Vous pourriez promettre à quelqu'un – qui, ensuite, ne vous déchargera pas de la promesse – que vous boirez la toxine demain après-midi. Ou vous pourriez encore signer un accord juridique vous obligeant à donner tous vos biens financiers (dont le million si vous le gagnez) au parti politique que vous aimez le moins, si vous ne la buvez pas. Vous pourriez même engager un tueur à gages pour vous éliminer si vous n'avalez pas la toxine. Vous serez donc sûr de passer une journée de misère, mais aussi de devenir riche.

Malheureusement, votre fille l'avocate, qui a lu le contrat attentivement, signale qu'un arrangement fondé sur de telles motivations extérieures est exclu, comme le sont d'autres astuces telles qu'engager un hypnotiseur pour implanter l'intention, oublier les principaux faits de la situation, et ainsi de suite (vous promettre *à vous-même* que vous boirez la toxine pourrait aider si vous étiez une des ces personnes singulières qui sont fières de ne jamais se désengager d'une promesse qu'elles se sont faites, quelles que soient les circonstances. Hélas, vous n'en êtes pas une.)

Obligé de vous rabattre sur vos propres moyens, vous essayez désespérément de vous convaincre que, malgré l'ordre chronologique, boire la toxine demain après-midi est une condition nécessaire pour empocher le million demain matin. Vous rappelant le problème de Newcomb, vous cherchez une preuve inductive que c'est ainsi, en espérant que les précédents bénéficiaires de l'offre du milliardaire gagnèrent le million lorsque, et seulement lorsque, ils burent la toxine. Mais hélas, votre neveu, un enquêteur privé, découvre que vous êtes le premier à recevoir l'offre (ou que de précédents gagnants burent moins souvent que de précédents perdants). Minuit approche maintenant à grands pas et dans un moment de panique vous essayez de susciter un acte de volonté (*summon up an act of will*), serrant vos dents et marmonnant « je boirai cette toxine, je boirai cette toxine » encore et encore.

Nous n'avons pas besoin de poursuivre ce récit de grands espoirs déçus (ou comblés) pour faire remarquer qu'il existe une énigme (*puzzle*) derrière tout cela. On vous demande d'avoir une simple intention pour réaliser un acte qui est bien en votre pouvoir. C'est le genre de chose que nous faisons tous de nombreuses fois par jour. Vous êtes doté d'une motivation irrésistible à le faire. Pourtant vous ne pouvez pas le faire (ou avez d'extrêmes difficultés à le faire) sans avoir recours à des astuces insolites comme vous faire hypnotiser, engager des

tueurs, etc. Vos difficultés ne découlent pas non plus d'une peur effrénée des conséquences négatives de l'acte en question – vous seriez en effet parfaitement disposé à subir les effets secondaires de la toxine pour gagner le million.

Deux points sous-tendent l'énigme (*puzzle*). Le premier concerne la nature des intentions. Si les intentions relevaient d'un travail intérieur ou d'ordres donnés à soi-même, vous n'auriez aucun mal à gagner le million. Vous auriez seulement besoin de garder un œil sur l'heure, et, à minuit, d'agir sur vous-même ou de vous donner un ordre à vous-même. De même, si les intentions étaient seulement des décisions, et les décisions des volitions sous le contrôle absolu de l'agent, il n'y aurait pas de problème. Or les intentions sont plutôt vues comme des dispositions à agir fondées sur des *raisons d'agir* – particularités de l'acte lui-même ou de ses possibles conséquences, que valorisent l'agent (déterminer la nature exacte du lien entre des intentions et les raisons sur lesquelles elles sont fondées est une tâche difficile et louable, mais qui ne doit pas nous retenir. Pour une explication semblable, dans l'ensemble, aux opinions présentées ici, voir « Intending » de Davidson, dans ses *Essays on Actions and Events*). Nous pouvons ainsi expliquer votre difficulté à gagner une fortune : vous ne pouvez pas avoir l'intention d'agir puisque vous n'avez pas de raison d'agir, du moins, vous avez de solides raisons de ne pas agir. Et vous n'avez (ou n'aurez lorsque le moment viendra) pas de raison de boire la toxine, et une très bonne raison de ne pas la boire, puisque cela vous rendra assez malade pendant une journée.

Cela nous amène à notre second point. Alors que vous n'avez pas de raison de boire la toxine, vous avez toutes les raisons (ou du moins un million de raisons) d'*avoir l'intention* de la boire. Et lorsque les raisons d'avoir une intention et les raisons d'agir divergent, comme elles le font ici, la confusion règne souvent. Car nous avons tendance à évaluer la rationalité de l'intention à la fois par rapport à ses conséquences et par rapport à la rationalité de l'action visée (*intented*). Ainsi, lorsque nous avons de bonnes raisons d'avoir l'intention d'agir, mais non d'agir, des critères d'évaluation contradictoires entrent en jeu et l'un d'eux doit céder : soit l'action rationnelle, soit l'intention rationnelle, soit, encore, des aspects de la rationalité même de l'agent (exemple : sa juste croyance que boire la toxine n'est pas nécessaire pour gagner le million).

J'ai fait les mêmes remarques dans un précédent article (« Some paradoxes of Deterrence », *Journal of philosophy*, Juin 1978), mais j'examinais alors un exemple qui mettait en jeu des intentions conditionnelles. L'énigme de la toxine élargit l'application de cette discussion, en montrant que ses conclusions peuvent s'appliquer à des cas qui mettent également en jeu des intentions inconditionnelles. Elle révèle aussi que les intentions sont

volontaires mais seulement en partie. On ne peut pas avoir l'intention de faire tout ce que l'on veut avoir l'intention de faire, pas plus qu'on ne peut croire tout ce qu'on veut croire. Si nos croyances sont contraintes par nos preuves, de même nos intentions sont contraintes par nos raisons d'agir.[1]

Traduction : Lucile Bouillant

[1] L'énigme examinée ici provient d'une conversation avec Tyler Burge, il y a quelques années, à propos de « Some Paradoxes of Deterrence ». J'en ai discuté utilement avec Paul Humphries, Rick O'Neil, et Virginia Warren, mais suis seul responsable de sa forme présente et des conclusions qui en sont tirées. Je suis reconnaissant à Doris Olin d'avoir suggéré un changement nécessaire dans un précédent brouillon.